AFRL-HE-WP-TR-2006-0028

AIR FORCE RESEARCH LABORATORY

HEAD MOUNTED ALERTING FOR URBAN OPERATIONS VIA TACTICAL INFORMATION MANAGEMENT SYSTEM

Susan Gottschlich
Robert Gray

Raytheon Company
1001 Boston Post Road
Marlborough MA 01752

Michael Daily
Ron Azuma
Youngkwan Cho

HRL Laboratories LLC
3011 Malibu Canyon Road
Malibu CA 90265

MARCH 2006

FINAL REPORT FOR 8 MARCH 2005 TO 28 FEBRURY 2006

Human Effectiveness Directorate
Warfighter Interface Division
Wright-Patterson AFB OH 45433

NOTICE

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2006-0028

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

FOR THE DIRECTOR

//signed//

MARIS M. VIKMANIS
Chief, Warfighter Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*: March 2006

2. REPORT TYPE: FINAL

3. DATES COVERED *(From - To)*: 8 Mar 2005–28 Feb 2006

4. TITLE AND SUBTITLE

Head Mounted Alerting for Urban Operations via Tactical Information Management System

5a. CONTRACT NUMBER

5b. GRANT NUMBER: FA8650-05-C-7231

5c. PROGRAM ELEMENT NUMBER: 62301E

6. AUTHOR(S)

Susan Gottschlich*, Robert Gray*, Michael Daily**, Ron Azuma**, and Youngkwan Cho**

5d. PROJECT NUMBER: 7184

5e. TASK NUMBER: 11

5f. WORK UNIT NUMBER: 36

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

* Raytheon Co., 1001 Boston Post Rd, Marlborough MA 01752
** HRL Labs LLC, 3011 Malibu Canyon Rd, Malibu CA 90265

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Materiel Command
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Wright-Patterson Air Force Base OH 45433-7022

10. SPONSOR/MONITOR'S ACRONYM(S): AFRL/HECV

11. SPONSOR/MONITOR'S REPORT NUMBER: AFRL-HE-WP-TR-2006-0028

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Cleared by SAF/PAX-06-0171, DOD-06-S-1312, AFMC-06-120, AFRL/WS-06-0799, 17 May 2006.

14. ABSTRACT

The United States military possesses unprecedented tactical alert generation capabilities but could quickly overwhelm a soldier conducting an urban operation with too much information. For this program, the authors investigated the use of a proof of concept Information Management Engine (IME) to allow a soldier to filter the information he receives, via a head mounted presentation system, through an intuitive training process. For the authors' purposes, the pieces of information that are sent to the soldier are referred to as 'alerts' and can be in the form of text, audio (speech), imagery, or streaming video. The objective is to provide a "peripheral awareness" capability that presents appropriate information via a head mounted see-around video display and an integrated earphone. Toward this end, the authors developed a prototype Tactical Alert Management System (TAMS) that uses the IME to determine if and how an alert should be presented to the user. The authors then developed a set of experiments to assess the military utility of the TAMS concept. Finally, the authors conducted an after action review and reported the results in this document.

15. SUBJECT TERMS Tactical alerting, head-mounted goggle display system, peripheral awareness, situation understanding, information management, incremental/active learning, multimodal user interfaces, ontologies, urban operations, asymmetric warfare, force multiplier

16. SECURITY CLASSIFICATION OF: UNCLASSIFIED			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Darrel G. Hopper
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	None	78	**19b. TELEPHONE NUMBER** *(include area code)*

This page intentionally left blank.

CONTENTS

LIST OF FIGURE CAPTIONS

LIST OF TABLE TITLES

FOREWORD

This contract FA8650-05-C-7231 was awarded under Defense Advanced Research Projects Agency (DARPA) Broad Agency Announcement (BAA) No. 04-31 entitled "Force Multipliers for Urban Area Operations," http://www.darpa.mil/baa/baa04-31.htm, on 8 March 2005 to Raytheon Company against their proposal number P-0431-100973 dated 5 October 2004 entitled "Head Mounted Alerting for Urban Operations," with government funding of $397,019 from DARPA for a 4 month technical effort (condensed from 6 months in proposal), hardware deliverables (including two sets of Micro-Optical binocular goggles) due by 7 October 2005, and a final report due in draft version by 7 August 2005 and approved version by 7 October 2005. The actual period of performance was extended. The final no cost time extension (NCTE), Modification P00003 dated 23 December 2005, made this a 10 month technical effort ending 31 December 2005 with a final report due in draft by 31 January 2006 and approved form by 30 March 2006. The total contract period was 8 March 2005 to 30 March 2006 (13 months comprising 10 technical effort and 3 final reporting). The formal final review was conducted on 25 August 2005 at the Air Force Research Laboratory (AFRL), Wright-Patterson AFB OH with a follow-up presentation at DARPA on 19 October 2005. There is no cost share (in cash or in kind) being provided by Raytheon for this effort.

This report has been formatted in accordance with a commercial standard, with tailoring from the AFRL Scientific Technical Information Office. This standard is as follows: "Scientific and Technical Reports—Elements, Organization, and Design," American National Standard ANSI/NISO Z39.18-1995 (NISO Press, Bethesda MD, 1995), which is available electronically via the following website address: http://www.wrs.afrl.af.mil/library/sti-pubh.htm

Measures of effectiveness (MOE) are critical to understanding the value of the present work. The contractor was asked to use the following standard definition in establishing MOEs: "Measure of Effectiveness – A quantifiable comparison of results obtained under specific external conditions and decisions. Examples include profit, quality, and customer satisfaction." Reference: Max's Wideman Comparative Glossary of Project Management Terms v3.1, http://maxwideman.com/pmglossary/PGM_M02.htm .

The Government Program Manager for this DARPA-funded AFRL-managed effort was Dr. Darrel G. Hopper of AFRL, who accomplished the technical review of this document.

PREFACE

The objective of this effort was to research, develop, and demonstrate an intelligent hands-free alerting mechanism driven by an IME that automatically correlates, prioritizes, and categorizes alerts based on user-developed rule set. The goal is to develop a minimally intrusive alerting system that shall allow a soldier to patrol and/or operate in an urban environment while being interrupted only by urgent and highly relevant Situational Awareness information. The end objective is to provide a "peripheral awareness" capability to the United States (US) soldier. Peripheral awareness means that, unlike the user defined operating picture (UDOP) approach to Situational Awareness, where the user defines the display based on his or her interests, peripheral awareness addresses stimulation of the user with Situational Awareness alerts – presenting information to the user that they were unable to anticipate.

New approaches to collaborative and all-echelon Command, Control, and Intelligence (C2I), along with automated intelligence analytical tools, are needed to support urban operations. Systems that can rapidly develop high-fidelity urban terrain maps with automated functionality and which can exhibit learning and tracking of past events are of interest. The goal is to create a collaborative environment that allows warfighters to see, understand, and interact with the urban battlespace in "real time." The urban approach requires tactical data at the level of the individual soldier that may be fed from local or remote networked sensors as well as new approaches for developing and predicting the effect of psychological operations. The C2I systems need to integrate unmanned capabilities with manned systems and provide support to the lowest level warfighter to include fire coordination for squad level indirect fire weapons.

The specific goals of this effort included: (a) the development of a proof-of-concept system design for a Head Mounted Alerting System (HMAS) for urban operations; (b) the development an experimentation campaign to assess this design; (c) the execution of the experimentation campaign; and (d) the accomplishment of an after action review (AAR) to analyze the results of the experiments.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support received from DARPA and program/contact management support from the AFRL. We thank Mr. Jeffrey Paul of DARPA, Mr. Bob Schulte of Solers Inc. (DARPA SETA Contractor), and Dr. Darrel G. Hopper of AFRL for their keen technical insights and guidance of this effort on behalf of the government.

This page intentionally left blank.

1. SUMMARY

This Head Mounted Alerting for Urban Operations program developed the proof of concept Head Mounted Alerting System (HMAS), illustrated in Figure 1.

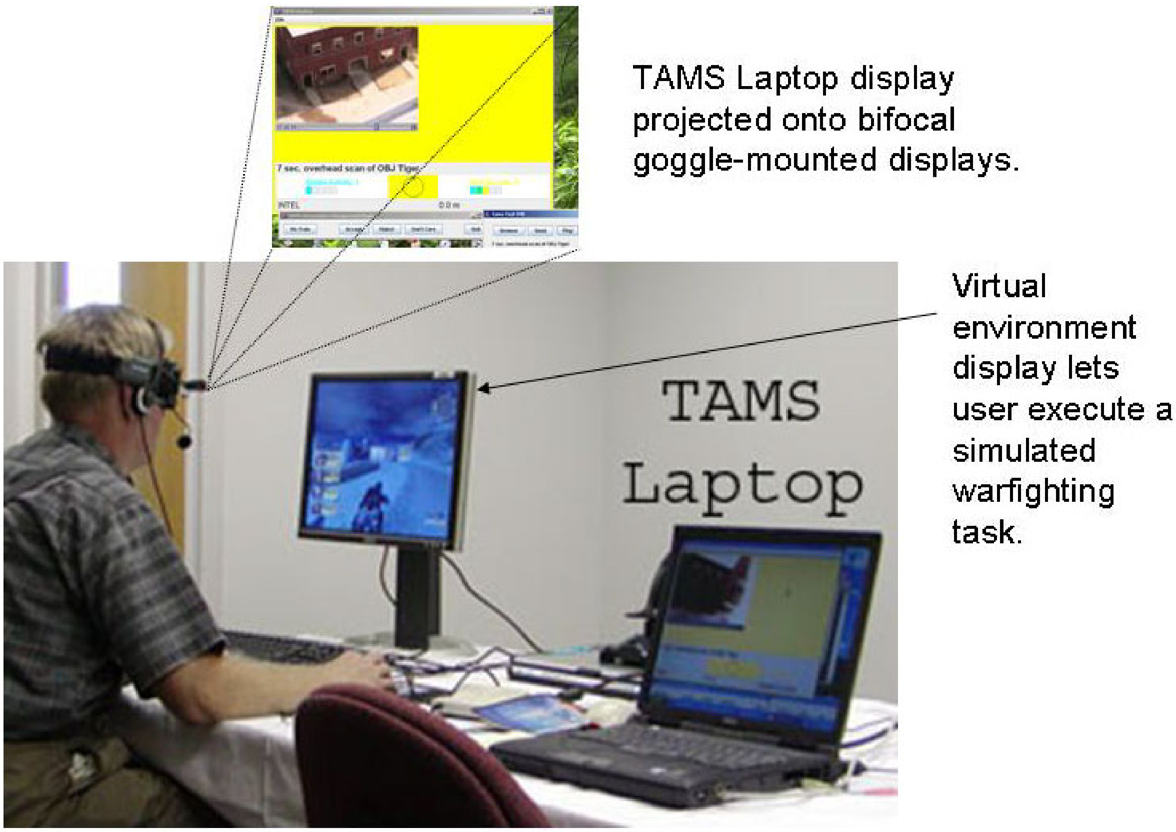


Figure 1: HMAS experimental setup.

The core innovation was an information management engine (IME) comprising three subengines for correlation and aggregation, user interaction, and display (visual and audio). The term “engine” here means software that encodes logic, algorithms, and data. This HMAS development, also known by its software name of Tactical Alerting Management System (TAMS), involved accomplishment of the following tasks:

- Researched and developed a proof of concept IME that “learned” how to filter alerts based on a novel supervised machine learning system that allowed a user to train the IME using a drastically reduced training set and a simple Accept/Reject interface
- Developed a TAMS prototype based on IME (see Figure 2)
- Designed and conducted an Experimentation Campaign to assess the design and military utility of the prototype. Wrote an After Action Review to document the results.

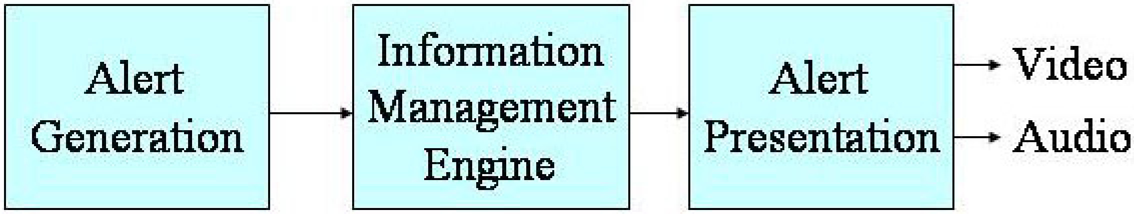


Figure 2: TAMS prototype software architecture.

The TAMS prototype involved a goggle-mounted see-around display and headphones, tethered to a laptop computer that presented tactical alerts to a user. The system was targeted for a non-technical user with understanding and possibly experience in urban operations and asymmetric warfare. For the purpose of the experimentation campaign, five different experiment configurations were created, four of which were used to immerse the user in a situation where they were required to perform a task and TAMS was used to present information, in the form of alerts, intended to assist the user in performing their task safely and effectively. The fifth configuration was created to allow the user to train the IME. A picture of the prototype system used during the experimentation campaign is depicted in Figure 1.

The alert generation module was created specifically to facilitate the individual experiments. It provided the ability to simply and precisely script the generation of alerts, to automatically perform data collection, and a manual mechanism for controlling the timing of the alerts sent to the user so that, if necessary, it was possible to synchronize the alerts with the task being performed by the user.

The IME addressed the core technical challenge in the TAMS prototype. As has been shown in this report, even for a relatively simple alert format, the multidimensional space of all possible alerts (i.e. alert space) is extremely large making it untenable for a commander or anyone else to develop alert management criteria by inspection. Furthermore, optimal alert management criteria must be specialized for each user since each user has a distinct mission in any urban operation. Allowing the user to develop individualized management criteria is the simplest approach, but it is unreasonable to expect each user to master a complicated user interface to do so.

Instead, a simple Accept/Reject/Don’t Care interface is provided to the user that lets him train the IME in a very intuitive fashion. Unfortunately, using such a simplified interface in a standard supervised learning system would require the user to consider a huge number of alerts that completely samples the alert space. An algorithm has therefore been developed whereby baseline management criteria is developed by inspection and sampling of the alert space.

The alert presentation module synthesized the video image displayed on the TAMS goggle-mounted see-around display and the audio played on the TAMS earphones. Three basic presentation modes were implemented – text-only, normal, and multi-modal. Simple rewiring allowed testing with two more presentation modes – audio-only and video only. These modes were required to support various experimentation objectives. Figure 3 depicts the alert presentation module displaying an image alert.

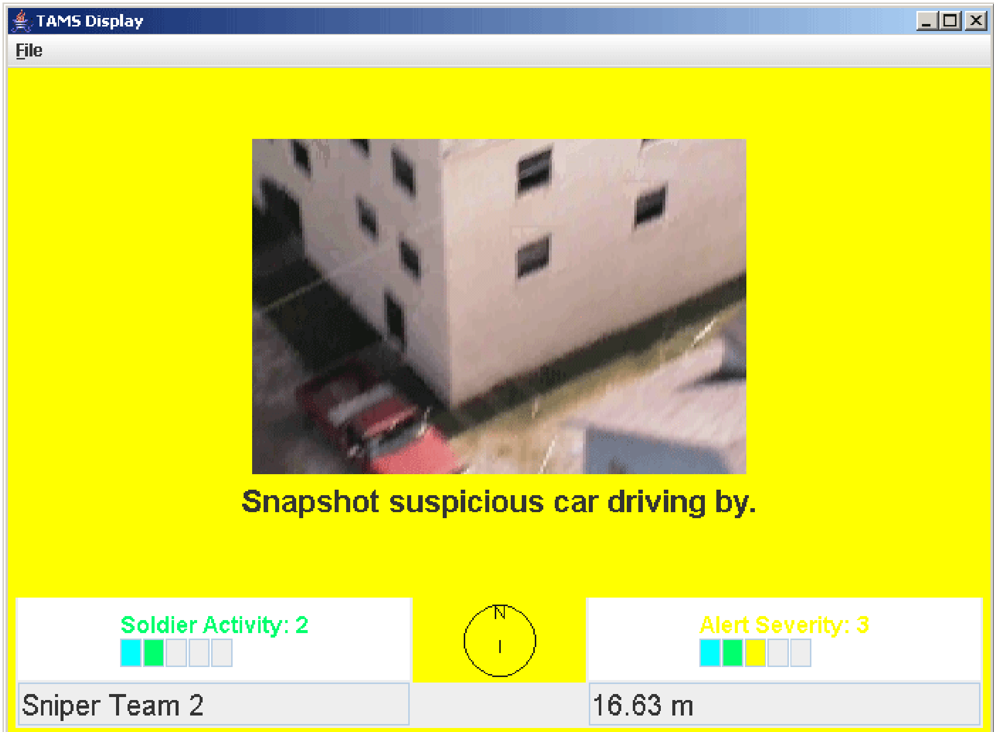


Figure 3: Example image alert synthesized by alert presentation module.

In order to develop an experimentation campaign, four qualitative Measures of Effectiveness (MOE's) were identified:

- MOE 1: Peripheral Awareness Enhancement.
- MOE 2: HMAS Interference with Situational Awareness.
- MOE 3: TAMS Training.
- MOE 4: Situation Understanding.

Based on these MOE's, measures of performance (MOP's) were defined to enable an assessment of each MOE. The five MOP's defined for MOE 1 are discussed in Section 3.6.1, the four MOP's identified for MOE 2 are defined in Section 3.6.2, the seven MOP's selected for MOE 3 are defined in Section 3.6.3, and the three MOP's defined for MOE 4 are addressed in Section 3.6.4.

Five experiment configurations were developed to assess the MOP’s as follows:

- Training Configuration – The Training Configuration involved only the TAMS prototype. Several individual training experiments were conducted during the campaign.

- Simple Configuration – A simple fictional web-site, “ganges.com”, was developed to provide a simple task for the user to perform while receiving TAMS alerts. The basic goal of this configuration was to provide a mechanism for measuring time saved, using TAMS alerts, when performing a simple task. The simple task was to place an order for a short list of items from this fictional web vendor. The alerts helped direct the user’s search through the ganges.com inventory.

- Situational versus Peripheral Awareness Configuration – A very repeatable task was developed to provide a mechanism to measure the user’s ability to react, in a timely fashion, to his surroundings and peripheral alerts being presented on TAMS. The Civil Emergency Reaction and Responder Training System (CERRTS) 3D (three dimensional) display was used to show the user a sequence of four simulated Unmanned Air Vehicle (UAV) flights around Norfolk harbor. An alert stream presenting emergency responders responding to a sniper incident in the harbor was presented on TAMS. The user called out “Norfolk” any time they saw the work “Norfolk” appear on the 3D display and “Sniper” any time they saw or heard the word sniper in an alert. These times were recorded and compared to a baseline.

- Video Game Configuration – A novice-level Operation Iraqi Freedom (OIF) inspired urban combat video game was purchased and partially scripted to allow an objective and repeatable mechanism to immerse the user in an urban combat-like situation and to collect objective measurements on red kills, blue lives saved, time required to complete mission, etc.

- Hostage–Rescue Configuration – Two hostage-rescue scenarios were implemented in CERRTS and the users were required to play the scenarios in order to assess their ability to respond correctly to their situation with and without the assistance of TAMS.

Each individual experiment involved using one of these configurations. Configuration parameters (e.g. performing task with or without TAMS) was adjusted to satisfy the objectives of the experiment.

Six different vignettes (scenarios) were developed to support these configurations:

- Training vignettes – Three complete hostage rescue vignettes were developed for use in training experiments. These vignettes were similar but involved different unit assignments and different unforeseen events.

- Sniper vignette – One complete emergency response to a sniper incident vignette was developed to use in the situational versus peripheral awareness configuration.

- Hostage rescue vignette – Two vignettes were developed to support the hostage rescue configuration. The first had some minor unexpected events, the second had a major unexpected event.

In addition, four different video game subtasks were partially scripted and a set of pre-scripted alerts were created for each video game subtask.

The experiments were performed over a three day period (18 – 20 July 2005). Subjective and objective data were collected and analyzed for each experimental configuration. The MOE results support an overall positive assessment of HMAS and its TAMS software.

Peripheral alerting generated by TAMS was enthusiastically received by the 12 subjects who wore HMAS during the experiments. Users quickly adapted to the technology, learned to employ its peripheral alerts to be more productive, and appreciated its potential for further development.

2. INTRODUCTION

The goal of this program has been to combine innovative technologies in a unique way to provide a minimally obtrusive peripheral awareness alerting mechanism to an operational user. Unlike the user defined operating picture (UDOP) approach to Situational Awareness where the user defines the display based on their interests, peripheral awareness addresses stimulation of the user with Situational Awareness alerts –presenting information to the user that they were unable to anticipate. The ideal peripheral awareness capability would present to the user all of the information they need but nothing extraneous. While fully implementing and fielding a peripheral awareness alerting solution requires several technical challenges to be overcome, the focus on this proof-of-concept phase has been on the intelligent information management of alerts presented to the user.

Peripheral awareness is achieved by providing the user with timely Situational Awareness alerts on a head-mounted see-around display illustrated in Figure 4 with integrated headphones. To keep the user from being overwhelmed with too much information, the IME analyzes every alert to determine if and how it should be presented to the user. The IME is based on an innovative combination of learning technologies that allows a user to “train” the IME by punishing or rewarding alerts that are presented to the user during a training session.

The video output of the IME is displayed on the goggle-mounted display (GMD) depicted in Figure 4. The GMD is an innovative new technology that Raytheon has invested in for application to the Driver’s Visual Enhancement domain. The GMD, in its current configuration, is a pair of goggles with a free-floating 600x800 pixel 18-color miniature monitor mounted under each eye.

Figure 4: Goggle-mounted binocular color SVGA see-around display.

Raytheon has recently committed new investment funds for technology improvements aimed at increasing the resolution and input capabilities of their head-mounted viewer product line and militarizing and hardening the packaging. The GMD offers a viewer that provides both a minimally intrusive head mount and an amazingly clear image.

Training is by example and counterexample. This paradigm results in a simple interface for the non-technical user to adjust the IME for their role, and to update their IME easily as their role evolves in reaction to changing conditions. We believe that the strength of the IME technology is in its simplicity, given that it is so easy for a user to interact with it, the user is far more likely to employ it accurately and often, thereby providing superior results to more complicated mechanisms.

Audio alerts and alert amplification are played on a set of earphones that the user wears in addition to the goggles. For the purposes of the proof-of-concept system, the goggles and earphones are tethered to a laptop that runs the TAMS software.

In order to evaluate and test the design and military utility of the TAMS prototype, this effort culminated in a three day experimentation campaign that assessed subjective and objective metrics of the TAMS prototype. Raytheon initially proposed using their CERRTS simulation to drive these experiments. However, given the aggressive goals levied on the experimentation campaign during the course of this program, Raytheon adopted its experimentation campaign to include four alternative experiment configurations, in addition to the basic Training Configuration, each involving a different computer application that was used to drive a subset of the experiments.

In developing the experimentation campaign, Raytheon collected input from both DARPA and AFRL during the kickoff meeting, a follow up teleconference, and the midterm integration event. In addition, Raytheon talked to former warfighters about the technology and what key aspects would be the most important to assess. In response to this input, four MOE's were developed for the TAMS prototype and several MOP's to assess each MOE. Based on these, a set of experiments to be used to assess each MOP and MOE was then developed.

Figure 5 depicts a screen capture of the TAMS output image prototype as displayed on the user's goggle. The prototype video output was simultaneously displayed on the laptop, from which this image was captured. The largest window is the alert display which will always be displayed on the goggles. The alert display application was developed by Raytheon. HRL assisted Raytheon in this development by providing subject matter expertise on interface mode ergonomics and related issues. Minor modifications were made to the layout of the alert display several times during the experimentation campaign to accommodate user feedback. The output image layout depicted in Figure 5 is the final version.

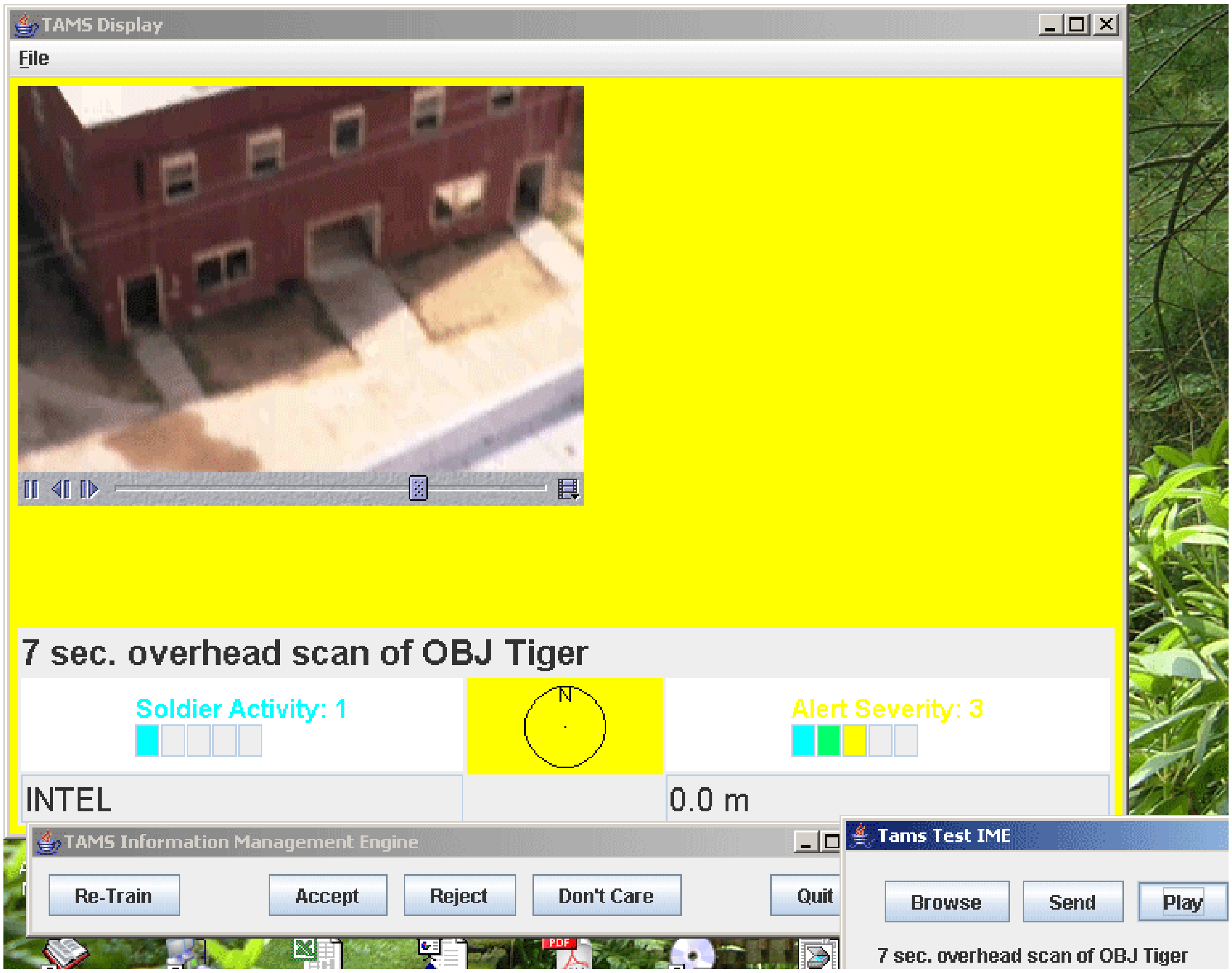


Figure 5: TAMS output image format.

The small window on the bottom left shows the IME graphical user interface (GUI). This offered the user the opportunity to accept, reject, or ignore (don't care) an alert. If the user did not react to an alert before a new one was displayed, then the IME took that to mean the same as clicking the "Don't Care" button. The user could click the "Re-train" button at any time, but it took several seconds for the IME to retrain its filter based on the user's input to the previous alerts, so this was only done at the end of a vignette during the experimentation campaign. The IME was developed by HRL Laboratories, LLC (HRL).

The small window on the bottom right is the alert generation utility that was developed to accommodate the four different experiment configurations that were used for the experiment campaign. This was initially developed as a simple testing mechanism, but as the complexity of the configurations increased, it was determined that passing all alerts through this utility during the experiments was the best way to insure robustness and repeatability during the experiments and also to automate some of the data collection required by the experiments. Raytheon developed this utility.

In order to guarantee the repeatability of the experiments, all alerts were generated and reviewed prior to the experiments. A set of alerts for each vignette was codified in a worksheet of an Excel file. Prerecorded audio, video, and images were stored in native format and linked into the respective alert fields in the Excel document. Each Excel worksheet was then saved as a tab delimited text file (there can be multiple worksheets in an Excel document). This document was read, line by line, by the alert generation utility, processed, and sent to the IME. The Pause/Play button on this utility could be clicked to temporarily pause and restart the sending of alerts. This was needed to allow for the manual synchronization the alerts for some of the experiment configurations. The user activity level could also be manually adjusted by this utility.

Four different alert modes were implemented: Text, Audio, Image, and Video. An amplification field on the alert was used to include a textual comment with all audio, image, and video alerts, which was displayed with each alert. For the most part, image alerts and video alerts were created from imagery and video either found on the web or captured from the software applications used during the experiments. Voice synthesis tools were used to create consistent voice-based audio alerts. It was found that synthesized voice alerts were less distracting than recordings of human voices.

Even for the simple training experiments where the user was doing nothing more than training the IME, detailed vignettes were developed and thoroughly reviewed. This was done because it was decided that it was extremely important for the user to be presented with a realistic set of alerts in order to make the best decisions possible during training. It also helped the user to think about how the technology might be used when evaluating its performance. Along these lines key locations in and around Baghdad were mapped out (and in the case of one vignette, Norfolk Harbor), roles and responsibilities for 5-10 units were defined (e.g. 1/A/1/75 Ranger Team, Sniper Team 1, 1/6 Infantry, etc.), and sets of alerts along a story line detailed in each vignette were created.

Given the aggressive schedule, limited scope of the program and unclassified nature, open source media and data consisting of images and video were collected and used to support tactical scenarios and alerts. The authenticity of the alerts was somewhat limited by this constraint.

3. METHODOLOGY AND RESULTS

The overarching objective of this contractual effort was to develop and perform experiments with a prototype for a hands free, unobtrusive peripheral awareness capability for use by an individual soldier or collaborative teams of soldiers operating in urban areas. For this contractual effort, Raytheon and HRL focused its investigation on the use of the IME for managing alerts. In particular, we concentrated on exploring the suitability of an interactive training mechanism, the lynchpin technology in the IME, to the peripheral awareness domain.

The term 'user' in this report refers to subjects who participated in the experiments. These subjects and their experiences are described in Table 1.

Table 1: Characterization of the subjects used in experiments.

Subject Role	Prior Military Experience	Urban Operations Experience	Discipline (at Raytheon)
Monday PM Test Subject	Navy	OIF (with Raytheon)	System Engineering
Monday PM Test Subject	Army	Unknown	System Engineering
Tuesday AM Test Subject	Army/Coast Guard	Boardings of presumed hostile ships.	System Engineering
Tuesday AM Test Subject	Navy	OIF (with CENTCOM)	System Administrator
Tuesday PM Test Subject	Army	OIF (with Raytheon)	System Engineering
Wednesday AM Test Subject	None	OIF (with Raytheon)	Quality Assurance
Wednesday AM Test Subject	None	None	System Engineering
Wednesday PM Test Subject	Navy	Navy	Quality Assurance
Video Game Test Subject	None	None	System Administrator
Video Game Test Subject	None	None	System Administrator
Vide Game Test Subject	Air Force	None	Chief Scientist

3.1 Concept of Operations (ConOps) Investigation

The objectives of this contract required us to establish a ConOps for the HMAS and the associated experiments. This section discusses in detail the methods and results achieved.

Raytheon developed the ConOps for TAMS by talking to former Special Forces soldiers and soldiers returning from Iraq, interfacing with DARPA and AFRL at two meetings and during one teleconference, talking to subject matter experts within the company working with warfighters on related programs, and by reviewing various documents and briefings from related programs.

Raytheon presented a preliminary ConOps at the Integration Event in May. This ConOps was adjusted based on AFRL and DARPA feedback. The primary adjustment was to focus on peripheral awareness and to not use TAMS for any standard Command and Control (C2) messages. The ConOps drove the development of alert messages needed to support the experiments.

Table 2 illustrates the ConOps as it applies to the hostage rescue vignette.

The basic unit of information in TAMS is the alert. Part of the ConOps investigation required that the structure of an alert be defined. Table 3 describes the definition agreed upon by Raytheon and HRL and used in the final experiments. In this table “Soldier” refers to the TAMS user. Note that the “Alert Text” field contains the alert payload. By payload is meant the actual content of the alert (e.g. a text string, a voice stream, an image, or a video clip). For the purposes of the proof of concept effort, this field is populated with a file descriptor accessible over the TAMS local area network to represent audio, image, and video payloads. In follow on work, the actual audio clip, image, or video clip may be contained in the payload as a binary large object (BLOB).

Table 4 shows some sample alerts used during training. The unutilized fields and the alert time field have been omitted to save space.

As the ConOps has evolved over the course of this program, it has been clear that TAMS needs to address the generation of alerts by members of the soldier’s unit, adjacent and nearby units, as the most important alerts will probably be generated by these soldiers and not at the headquarters.

Table 2: TAMS ConOps applied to hostage rescue vignette**.**

Vignette Step	How Now?	How with TAMS?
BCT Intelligence Officer disseminates Intelligence Summary indicating increased insurgent/fighter interest in initiating an attack. 2BCT/1AD on routine patrol of Al Rashid and AL Mansur are informed an American newswoman has been taken hostage.	Traditional C2 messaging/procedures. Would likely take a meeting to discuss hostage and disseminate her picture.	Via TAMS text, audio, and/or imagery. Hostage picture can be disseminated using TAMS so that everyone can immediately be on the lookout for insurgents moving her.
1/6 Infantry begins search of area to gather information that may confirm or deny informant report.	Hand signals and radios.	Soldiers exchange imagery, messages via TAMS.
BCT S-2 works with higher headquarters to confirm or deny report.	Traditional C2 messaging.	Same as now.
2 BCT Commander is given mission of securing area while Special Forces and Ranger unit prepare to attempt a rescue while waiting for direction from higher headquarters.	Voice comms.	Soldiers from each unit stay in touch using TAMS. Predator feeds and imagery from SOC available via TAMS.
2/6 Infantry mans check-points along Highway 10 in vicinity of Al Rashid in the south and the lines of communication leading into Al Rashid and Al Mansur from the north, east, & west.	Voice comms and hand signals.	Team exchanges images and messages using TAMS. Predator feeds and other imagery available via TAMS.
Special Forces and Ranger unit receive go ahead to attempt a rescue mission.	Voice comms.	Via TAMS.
Forces carry out a dress rehearsal in Green Zone	Voice comms.	TAMS available to monitor situation during rehearsal.
Rescue mission is executed.	Situational Awareness from UAV, Blue Force Tracker (BFT). Voice comms from HQ.	Images/messages exchanged between snipers, rangers, and SF teams using TAMS. Predator feeds and other imagery available via TAMS.

Table 3: Alert definition

Field	Description
Mode	Text, Audio, Image, or Video
Type	Hardcoded to "Plain Text" to maintain place holder, not utilized in final TAMS prototype.
Soldier Activity Level (SAL)	Value 1-5 describing the soldier's activity level. 1 means not busy, 5 means extremely busy.
Alert Severity Level (ASL)	Value S/1-S/5. S/1 means informational, S/5 means extremely severe.
Capability	Hardcoded to "Comms" to maintain placeholder, not utilized in final TAMS prototype.
Responsibility	Which one of the seven Battlefield Operating Systems (BOS's) to which the soldier belongs. Hardcoded to "MANUEVER" upon direction from DARPA to focus on that BOS.
Alert Location	Geographic coordinate describing the location associated with the alert.
Soldier Location	Geographic coordinate describing soldier's location.
Alert Unit	Unit responsible for generating the alert.
Alert Time	Time that alert was generated.
Soldier Unit	Unit to which soldier belongs.
Alert Text	Payload of the alert. For the purposes of this phase, this is a text string for text messages, or a universal resource locator (URL) for audio, image, or video.
Alert amplification (amp)	Text amplification string associated with every type of alert other than text.

Table 4: Sample alerts from training vignette**.**

Mode	SAL	ASL	Alert Loc	Sol Loc	Alert Unit	Sol Unit	Alert Text	Alert Amp
Image	5	S/5	BE	BE	Sniper Team 1	ODA Team 2	\alertMedia\moutArea\odaTeamOnRoof.jpg	ODA Team 1 landing on roof of OBJ Tiger.
Text	5	S/4	OP 2	BE	Sniper Team 2	ODA Team 2	Armed man from 2nd floor neutralized.	
Video	2	S/2	CP 9	BE	2/6 Infantry	ODA Team 2	\alertMedia\baghdad\erraticHelo.avi	Helo acting erratically.
Audio	5	S/5	OP 2	BE	Sniper Team 2	ODA Team 2	\alertMedia\moutArea\movementInside.wav	Movement inside seems headed for back of building.

3.2 Information Management Engine

The objectives of this contract required HRL to design and develop the IME software comprising three subengines to accomplish alert correlation and aggregation, user interaction, and display. The IME is a supervised learning system that creates a decision engine based on a small set of tertiary inputs collected from a user during a training phase. During the training phase, the IME presents a sequence of alerts to the user, and the user responds to each alert by pressing either the “accept,” “reject,” or “Don’t Care” button on the IME user interface. If the user does not respond before the next alert is presented, the response is recorded as “Don’t Care.” Based on these responses, the IME develops a look up table representing the user’s preferences by running an incremental multi-class support vector machine classifier, this lookup table being the basic mechanism used by the decision engine to determine the presentation mode of an alert.

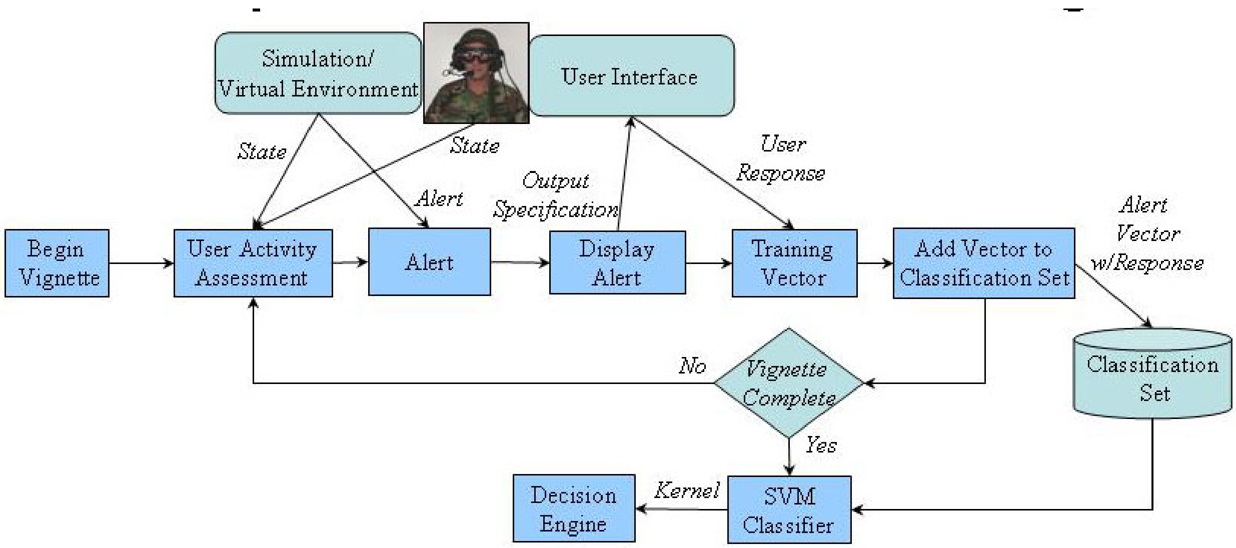
Three variations of this system were developed: a supervised batch training system, a supervised incremental training system, and an unsupervised training system. For the final experiments, the supervised batch and incremental training systems were merged into a single system that, in response to a user request, can update the classifier at any point during the training session.

Figure 6 shows the system architecture for these variations. The supervised batch training system was designed to accommodate user input while a specific simulated vignette was played out. On the other hand, the supervised incremental training version was designed to enable the system to update the classifier as each new user response was added to the classification set. In merging the two, a mechanism to allow the user to decide when to perform incremental updates during a simulated vignette was provided.

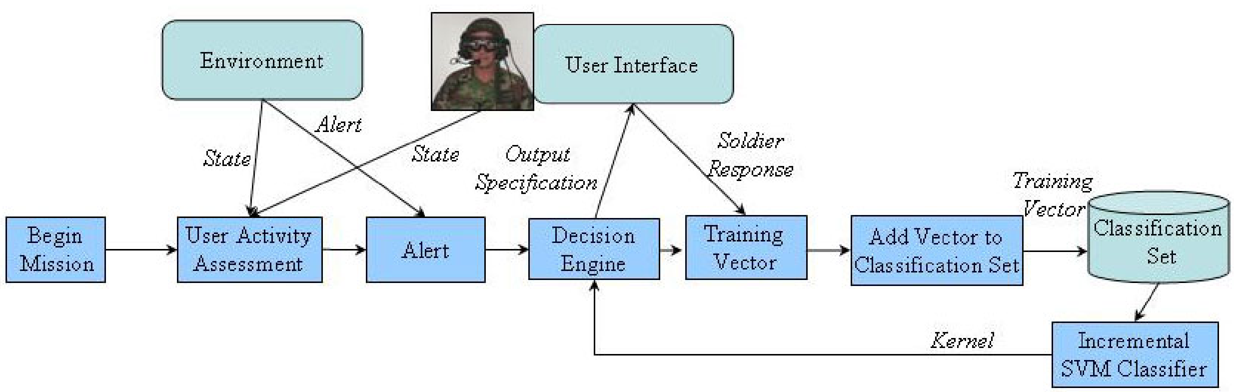
The unsupervised training system was designed to take as input a specification of the user’s preferences in rule form and construct a complete table of all possible alert input values and output preferences. This form was used primarily to measure the key performance metric for the IME, the learning rate.

3.2.1 Interface Ontology

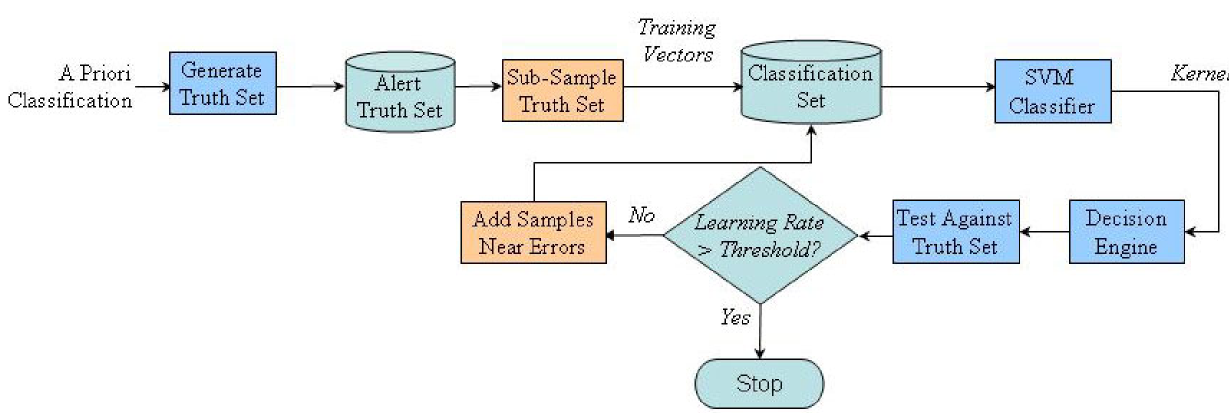
The interface ontology was developed using three spirals, each increasing the sophistication of capability over the last spiral. In the first spiral, the focus was on the development of an ontology that could support IME decisions for either display or no-display (see Figure 7). This enabled the focus to be on the prioritization of alert input values and the classifier method. In the second spiral, the ontology was generalized to support output modes of audio, text, imagery, and video in addition to no-display (see Figure 7). The final development spiral used specific heuristic constraints on the possible output modes (see Figure 8). Input modes could be transformed to a more basic output mode such that video -> imagery -> text -> audio (described below).



(a)



(b)



(c)

Figure 6: Training architectures – interactive vignette (top), truth set (middle), incremental (bottom).

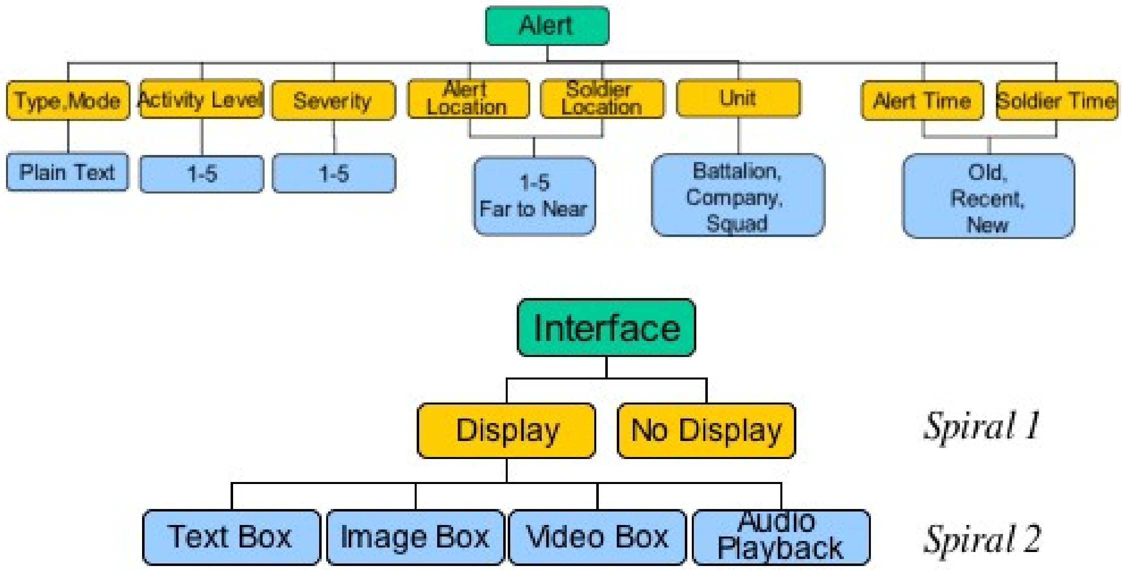


Figure 7: Ontology for Spiral 1 and Spiral 2 of IME.

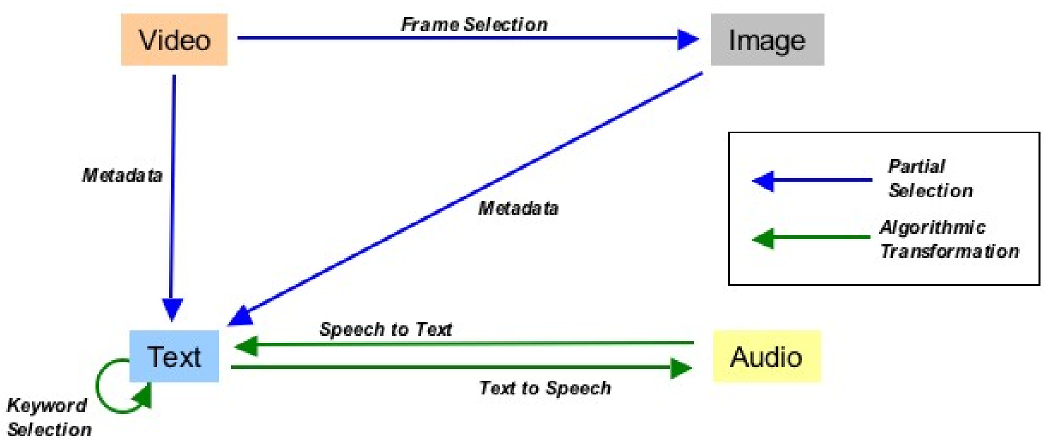


Figure 8: Selection or transformation of an input to an output mode.

An alert is input to the IME with the following fields, referred to as dimensions (the IME ignores some alert fields and combines others):

- Severity level (1 to 5, with 5 being high)
- Soldier activity level (1 to 5, with 5 being high activity)
- Relative Location (1 to 5, where 5 is close and 1 is far)
- Unit (1 to 5, 1 = lowest priority and 5 = highest priority)
- Display mode (1 = none, 2 = audio, 3 = text, 4 = image, 5 = video)

In addition, the unit and alert locations were defined relative to the receiver of the alert, requiring the alert values to be computed for that receiver based on a pre-defined policy. For the purposes of the experiments, the policy was defined as shown in Figure 9.

Using the above five dimensions, each with five discrete values, there are $5^5 = 3125$ distinguishable inputs, each with an output value of +1 (yes) or -1 (no) indicating the user's preference for display. The training and "truth set" formats are very similar, and differ only in the number of values stored. Each line has the following format

classification 1: activity 2: severity 3: location 4: unit 5: mode

So, for example, two lines might be as follows: -1 1:5 2:1 3:3 4:2 5:2
+1 1:1 2:5 3:4 4:4 5:4

which mean that the user says "reject" for an alert with activity=5, severity=1, location=3, unit=2, mode = audio, while the second line states that the user says "accept" for an alert with activity=1, severity=5, location=4, unit=4, mode = video.

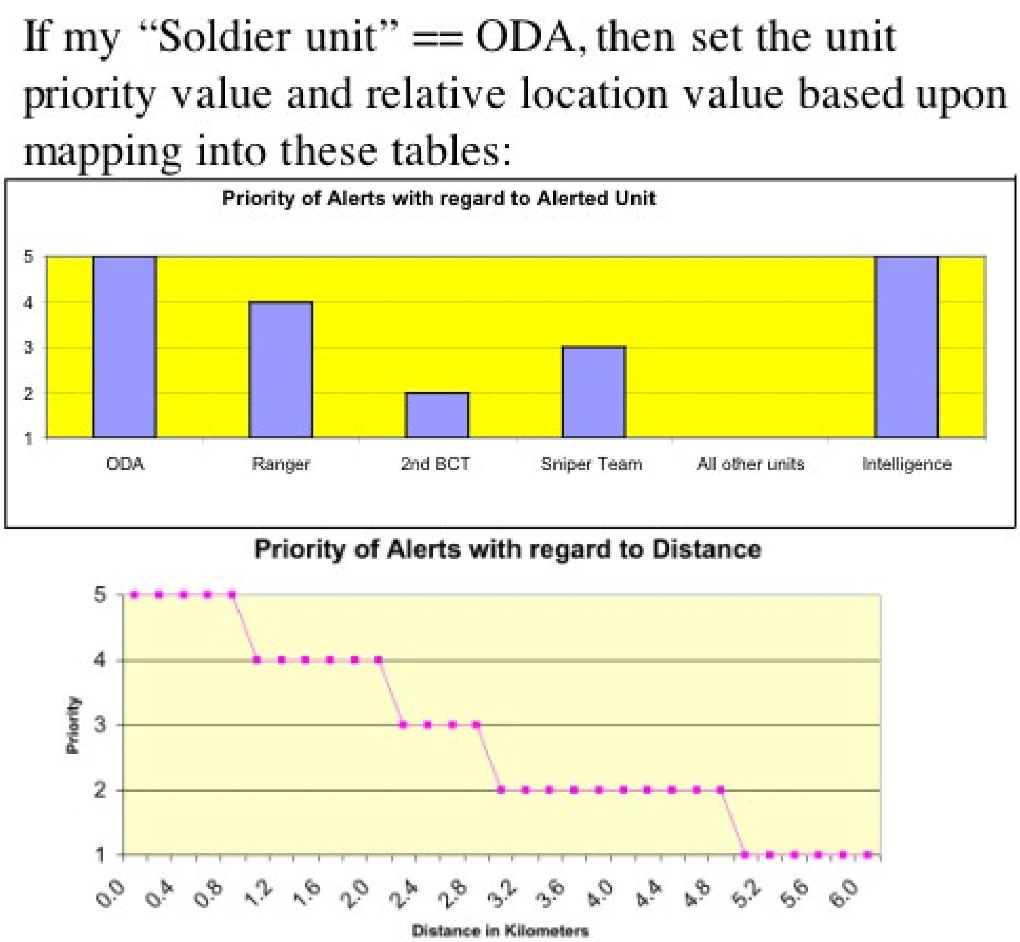


Figure 9: Unit and location values are computed relative to receiver (soldier unit ODA).

To store the results of the classifier, a truth table is used. The truth table has 3125 lines, containing all possible combinations of the five input dimensions. For each of those, the table states whether that alert has a yes or a no value. The IME decision engine reads the entire truth table and stores it into a lookup table, which is used to execute the decision logic for a particular alert. When an alert arrives, the IME looks up the truth table entry with those input values. If the result is +1, then the system shows the alert in the specified output mode. Otherwise, the lower-numbered display modes are examined, in order, to see if those have a positive output mode (+1). The output mode is specified as the first display mode encountered with a positive output mode. If none of the display modes have a positive output mode, the output mode is converted to "Don't Display". For example, if the IME receives an alert vector of the form {-1 1:3 2:4 3:3 4:4 5:5} (i.e., severity=3, activity=4, location=3, unit=4, mode = 5 (video)), since it is scored –1, the algorithm first checks the lookup table value for severity=3, activity=4, location=3, unit=4, mode = 4 and find it has an output mode of –1, so next the algorithm checks the lookup table value for severity=3, activity=4, location=3, unit=4, mode = 3 (text). This vector has a defined output = +1, so the output mode of this alert is changed from video to text. Therefore the TAMS alert presentation software will display alert amplification field rather than video in alert payload.

3.2.2 Learning Classifiers

The IME uses a multi-class adaptation of the support vector machine classifier to generate a decision surface. Support vector machines (SVM's) are "large margin" classifiers that attempt to construct a decision surface that maximizes the distance from the training set to the decision surface. The samples that are closest to the decision surface are called support vectors. SVM's have strong theoretical foundations and have been used successfully in a variety of real world tasks. In their simplest form, SVM decision surfaces are hyperplanes that separate the training data by a maximal margin such that all vectors lying on one side of the hyperplane (in a binary classification setting) are labeled as a –1 and all vectors lying on the other side are labeled +1 (Figure 10). The decision surface, or kernel, may be defined using a variety of different functions including polynomial approximations, radial basis functions, or user selected kernels. We used the freely available "SVM-Light" implementation of SVM's and modified it for multi-class use and incremental updates.[1] [2]

3.2.3 Historical Reference Models

In a fully developed TAMS/IME system, it is envisioned that a set of case histories (historical reference models) that represent the various user preference characteristics for choice of display. A historical reference model is a set of decision surfaces for several users that capture essential differences or bias in preferences for alert display. These models may change with different vignettes or other conditions. For example, Figure 11 shows single dimensional changes with a bias toward no display. These case histories would be used to initialize the IME prior to training, as described in the experiments section.

---

[1] Vapnik, V., *Statistical Learning Theory,* New York: John Wiley & Sons, 1998.

[2] Burges, C., "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* 2:121-167, 1998.

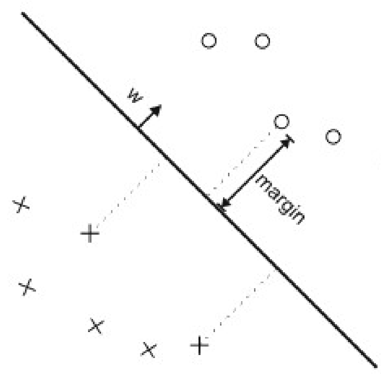


Figure 10: SVM's classify training samples by maximizing the margin between the training set and the decision surface.

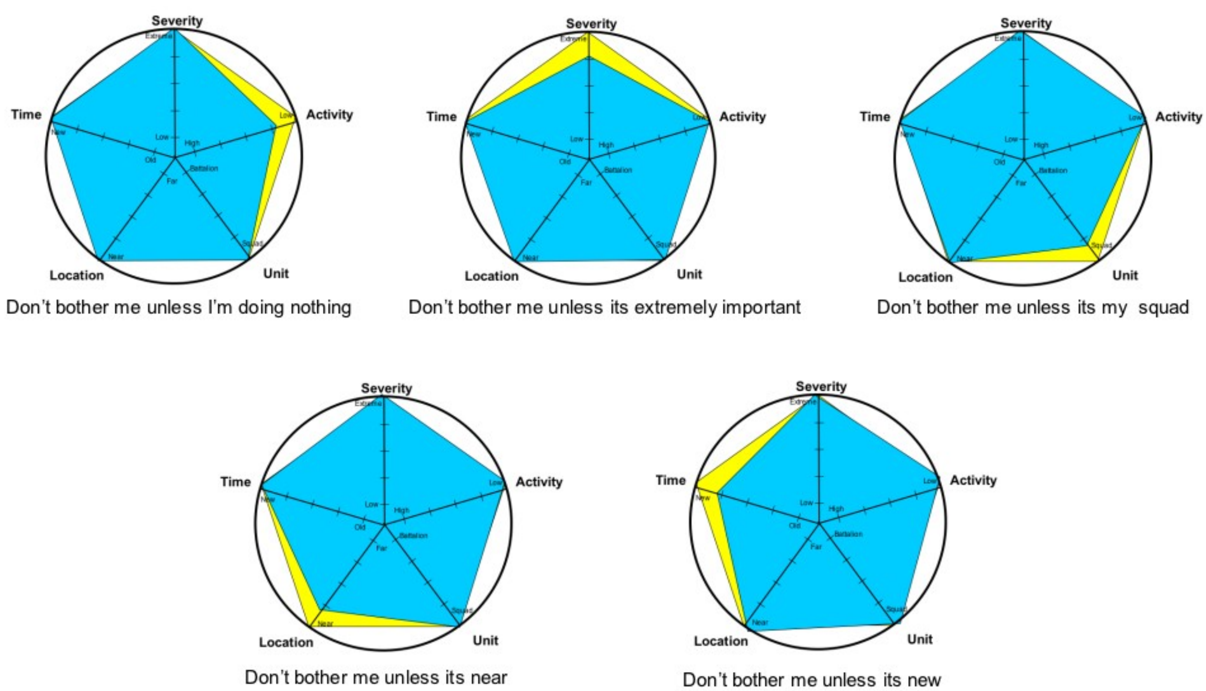


Figure 11: Historical reference models that vary in one dimension of the input vector and their interpretation. Blue-dark regions correspond to no display decision; yellow-light, display.

3.3 Tactical Alert Management System

The objectives of this contract required Raytheon and HRL to design and develop a prototype tactical alert management system. The prototype was comprised of four components: the alert generation module, the IME, the alert presentation module, and the wearable hardware (goggle

and earphones). For the purposes of the experimentation campaign[3], the TAMS software modules were hosted on a low end laptop, but could be run on almost any platform.

3.3.1 TAMS Alert Generation Module

The alert generation module was originally developed as a test tool to facilitate early integration between CERRTS and the IME. However, as the scope and ConOps of the experimentation campaign evolved, it was determined that the test tool should be expanded into the alert generation module which helped to make the experiments far more robust, predictable, automatable, and quantifiable. Furthermore, it provided an easy mechanism to experiment with preprocessing of alerts. For instance, during the course of this contract, Raytheon and HRL experimented with keyword extraction as an input dimension for the IME. The alert generation module performed the keyword extraction, and forwarded the keywords, along with the alert, to the IME. It was decided not to use keywords for an input dimension for the final prototype but may revisit their use at a later time.

The alert generation module did allow for the specification of “symbolic” locations in alert messages, such as “CP 9” and “OBJ Tiger”, which were then replaced with full coordinates extracted from a look up table. It also allowed timestamps to be specified in terms of a differential number of seconds since the last alert was sent. Since, in order to accommodate the revised ConOps, all of the alerts needed to be created manually, rather than automatically by CERRTS, these two specification mechanisms facilitated the manual creation of an alert, making the whole process less cumbersome and error-prone.

In order to support accurate and automated data collection, it was possible to append a keyword to the alert type field of any alert. Whenever the alert generation module came across such an alert, it stripped off the keyword and recorded it, and the current timestamp, in a log file. This mechanism was used to support data collection for some of the experiments.

The alert generation module was also used to synchronize the sending of alerts with the user activity during the experiment for experiment configurations where there was no other way to directly link the two. For instance, this method was used for the Video Game Configuration.

The alert generation module was an artifact of the prototype. It is envisioned that it will be completely replaced with an alert distribution module in a future phase. For these experiments, alerts were generated for each user individually, but in an end use, alerts would be generated by non-TAMS applications and it would be necessary to multicast these alert to one or more TAMS users on the network.

3.3.2 TAMS Alert Presentation Module

The TAMS alert presentation module presented the alert to the user along with the pertinent attributes that were used by the IME training algorithm to train the filter. These parameters are

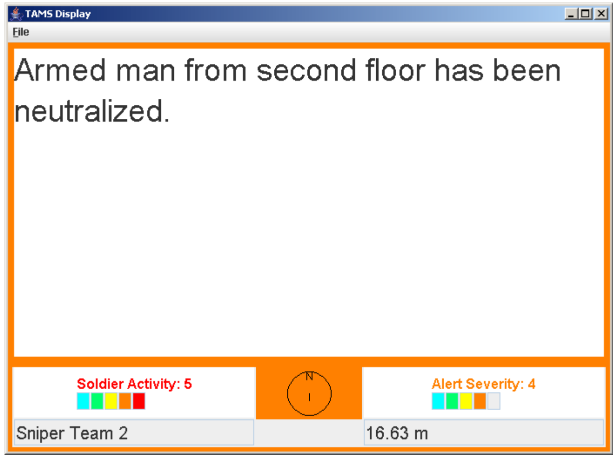
[3] Alberts, David S., Richard E. Hayes, John E. Kirzl, Dennis K. Leedom, and Daniel T. Maxwell, *The Code of Best Practice for Experimentation,* Washington, DC: CCRP Publication Series, 2002.

location (presented relative to the user, via distance and direction), unit generating the alert, alert severity, user activity level, and (implicitly) type of alert.

Text, image and video alert formats are illustrated in Figure 12. Note that an audio alert looks like a text alert. The text displayed on an audio alert is taken from the “amplification” field of the alert. Similarly, the text displayed under the image and video alerts is also taken from the alert amplification field if it exists.

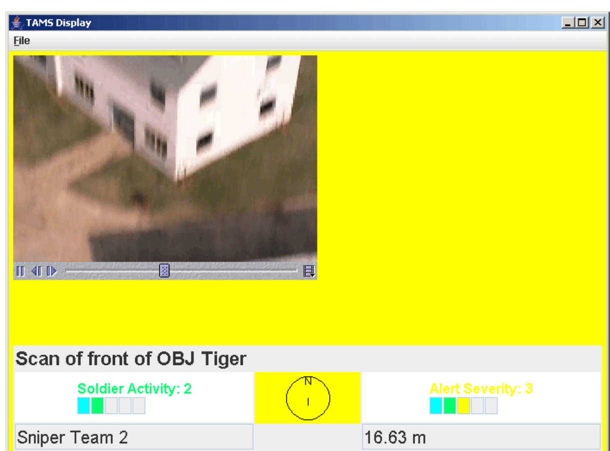
The alert presentation module was modified several times during the 3-day experimentation campaign to react to suggestions being made by the users. While an improvement over time of the users’ ability to understand and react to alerts was observed, there is no quantitative MOP to support this observation.

During the course of the experiments, the users made some very good suggestions that we simply could not implement while the experiments were being carried out but we will look at them in the future. Table 5 summarizes some of these comments.



Text Alert

Image Alert



Video Alert

Figure 12: Text, image, and video alert format examples.

Table 5: User comments regarding alert presentation

Use symbology, overlaid on concentric rings, to signify distance, direction, and subject of alert.
Make the border of the viewer a higher order color to make the “see-around” easier to focus around and back to.
When using text only, I was unable to function as well as the text with audio because I was not as aware of a change in the alerts.
Images are important, they need to be displayed. (See Note 1).
Background color too obtrusive
[TAMS would be better if it had] 3 dimensional video.
Make peripheral indicator more prominent on screen
Use more symbology less words
Use uniform font size. (See Note 2).

Note 1: The IME ignored the alert, passed the alert, or transformed the alert to a simpler mode. This user is reacting to the transformation of less important image alerts to text alerts. We concur with his statement – in some cases this transformation produced confusing results.
Note 2: Text alerts were displayed in a larger font size than the amplification text associated with the other alert types.

Three alert presentation modes were implemented to support our experiments:

- Text-only – Every alert, regardless of original type, was converted to a text alert by copying the alert “amplification” field into the text field and presenting it with no accompanying audio.

- Normal – Each alert was presented in its original form. If amplifying text was specified, it was displayed under the base alert (in the case of image and video) or in the same way as a text alert (in the case of audio).

- Multi-modal alert – A notification sound was played with each new alerts and a text-to-speech module was used to read the accompanying text (in the case of text, image, and video) while the alert was otherwise presented as described above. The notification sound was modified during the course of the experiments from a very unobtrusive sound to a more pronounced sound. It was observed that the users who were notified by the more pronounced sound were generally more aware of each new alert but we did not collect against an MOP to support this observation.

A video-only alert mode (take the earphones off the user) and an audio-only mode (take goggle-mounted display off user and use multi-modal alert mode in presentation module so that all alert amplification and text alerts are read to user) was also informally achieved during the experiments.

As the quantitative results presented have born out, it was found that providing the user with an audio cue every time a new alert appeared allowed the user to react more appropriately to both situations in front of them and to alerts. While there were concerns that audio might be distracting, the users found that having to look back and forth between the scene in front of them and their goggle display was more distracting than receiving an audio cue.

The TAMS alert presentation module was meant to only support the proof of concept experiments. However, since it was the primary user interface between the user and TAMS it was decided that it was important to make as good as possible to avoid distracting the user.

The TAMS alert presentation module was written exclusively in Java and utilizes the Java Media Framework to play audio and video and the FreeTTS open source Java speech synthesizer to implement text to speech amplification[4]. It was found that the use of automated text-to-speech synthesis produced less distracting audio tracks than recording a (familiar) human reading the same text.

No attempt was made to normalize the size of images or video. In the future, normalizing these will make it easier to stretch the alert field over a great portion of the TAMS display so that the alert maximizes its use of display real estate.

3.3.3 TAMS Wearable Presentation Hardware

The TAMS wearable hardware consisted of a pair of bifocal goggle-mounted see-around 800x600 color displays and a low-cost set of PC headphones with integrated microphones (microphones were not used for the experiment). For the purposes of the experimentation campaign, the wearable hardware was tethered to the TAMS laptop.

For the training experiments, where no other task other than the training needed to be performed by the user, experiments were performed with using a regular computer monitor and PC speakers rather than the display. No appreciable difference in the results was seen.

A minor problem fitting the goggles over some of the users' glasses was found. Whether they opted to take their glasses off or fit the goggles over (pressing the glasses against the head in a somewhat unnatural fashion), it was noticed that these users had a harder time seeing the alerts. It is clear that future generations of the see-around displays need to accommodate a variety of eyewear.

Future generations of TAMS need to move toward a less obtrusive, untethered presentation system. It is hoped that ongoing work at Raytheon on the Tank Urban Survivability Kit (TUSK) program and anticipated future programs can be heavily leveraged to develop a low power, modular, ruggedized audio/video (AV) system that is equipped with wireless communications capabilities.

[4] See http://freetts.sourceforge.net/docs/index.php.

Raytheon has recently won two contracts from the TUSK program that aims to deliver add-on kits to M1A2 Abrams tanks where goggle mounted displays will be delivering as part of the thermal weapon site package included with this kit. Under this contract, Raytheon has delivered two systems to Aberdeen and will be delivering two more soon. Aberdeen is scheduled to begin testing these systems along with other parts of the kit the week of 22 August. If the testing is successful, we expect a Low Rate Initial Production (LRIP) contract to be let no later than December. The end user for these systems is the operational mounted soldier.

The 82nd Airborne at Ft. Bragg has issued an Operational Needs Statement (ONS), attached in Appendix E of this document that calls for the use of a goggle mounted display to a thermal weapon site (TWS) on an M4 Carbine assault rifle. This ONS has been staffed by the Department of the Army to the U.S. Army Training and Doctrine Command (TRADOC) to perform a Doctrine, Organization, Training, Materiel, Leadership & Education, Personnel, and Facilities (DOTMLPF) analysis of this use. Raytheon anticipates a favorable recommendation from TRADOC which will likely result in a contract that will ultimately deliver goggle mounted displays systems to the Army to be used by operational dismounted soldiers.

Figure 13 depicts a conceptual prototype of the delivered system. The see-around militarized displays have become DX rated in the Defense Priorities and Allocation System, which is evidence of the support they are gaining with operational military units.

3.4 Experiment Design and Development

In order to meet the objectives of this contract Raytheon was required to design and develop experiments. The purpose of these experiments was to assess the TAMS design and military utility. Raytheon and HRL interacted with DARPA, AFRL, as well as with former warfighters in order to develop a short list of qualitative MOE's on which to base the campaign. These MOE's were then used to develop a set of quantitative MOP's that could be directly measured by the experimentation campaign. Based on the draft MOE's and MOP's five experiment configurations were defined which are detailed in Section 3.4.1. Based on these configurations, a schedule of experiments was then developed that is discussed in Section 3.4.2.

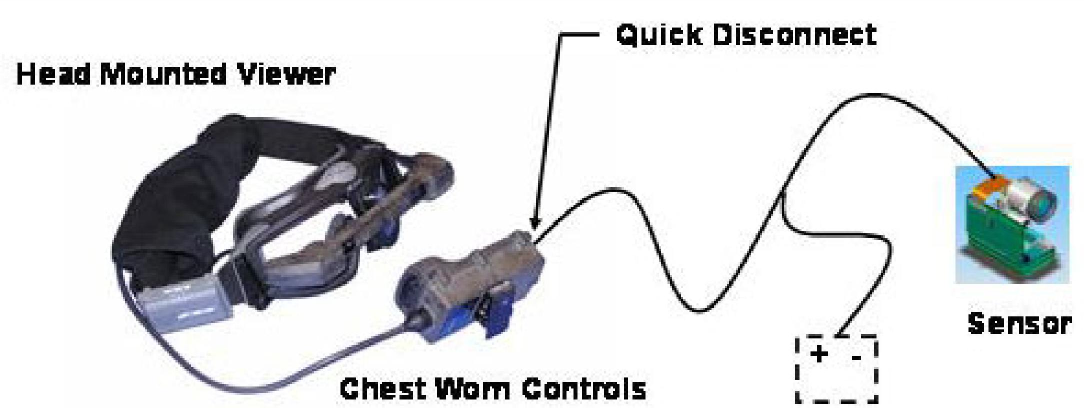


Figure 13: Conceptual prototype of control system for TUSK contract. A wireless system is being considered in addition to the tethered system depicted**.**

3.4.1 Experiment Configurations

In all, we developed five different experiment configurations. The selection of the configurations was driven by the needs of the experiment and the ease of implementation. At the beginning of the contract, only two configurations were planned – a Training Configuration and a Hostage-Rescue Configuration. However, as the ConOps of the TAMS prototype evolved and the needs of the experiment evolved, it was felt necessary to introduce other configurations to more accurately and objectively collect data against some of the MOP's.

3.4.1.1 Training Configuration

The Training Configuration was the simplest experiment configuration that was developed because it required only the TAMS prototype. The goal of this experiment configuration was two-fold: to assist in evaluating the TAMS Training MOE and also to train the TAMS prototype for use in other experiments.

This experiment configuration involved the complete TAMS prototype. The user was told to view the alerts and use the IME GUI to record their preference for each alert. If they did not react to an alert, it was recorded as "Don't Care." Initially, the IME GUI was much larger than the final configuration (see Figure 5), the GUI was reorganized during the experiments to allow for more screen real estate for the TAMS alert presentation GUI.

To support this configuration, it was decided that the training would be most effective if realistic vignettes on which to base the training alerts were developed. Along these lines, three different vignettes were developed that were relatively similar but that involved different unexpected events: e.g. ODA Team helicopter crashing, kidnappers in hostage building shooting at "friendlies", etc. There was also some variation on how some of the units (1/A/1/75, 1/6 Infantry and 2/6 Infantry, 1-35 Armor) were used. In all ten units were used in the vignettes: ODA Team 1, ODA Team 2, Sniper Team 1, Sniper Team 2, 1/6 Infantry, 2/6 Infantry, 1/A/1/75 Rangers, 2 BCT, Intel, and 1-35 Armor.

Over 150 alerts were created for each vignette. Text, audio, image, and video alerts were combined in a more or less even number and these alerts were interleaved as evenly as possible without compromising on realism. Audio alerts were prerecorded using a few different speech synthesis packages. Alert severity and soldier activity level were hard-coded based on the best assessment of what the soldier would be doing in the underlying vignette.

The alerts were spaced between 5 and 15 seconds apart based on the amount of time required to "view" the alert. For instance, the alert following a twelve second video alert may be sent 15 seconds after the video alert was first displayed. Data collection was built into the IME so that it could automatically collect data on the training process.

Generally speaking, experiments based on this configuration proved to be easy to run, as it was almost fully automated, and useful for later experiments. This configuration served a very useful purpose for the experimentation campaign – it essentially provided the user with a hands-on

interactive demonstration of the technology which helped them better understand the technology and its potential uses.

3.4.1.2 Simple Configuration

Raytheon took the action to develop a Simple Configuration during the Integration Event meeting. The notion was to give the user a simple task to complete that they would be able to complete faster if they had access to “alert” information.

Along these lines we developed a very simple “web site” that allowed users to order books, CD’s and DVD’s from a fictional web vendor called “ganges.com.” Two completely different inventories were developed for this site so that experiments based on it could be run twice before the user became familiar with the inventory. To minimize effort, the site was written in html with a small amount of JavaScript to perform tasks like totaling up the order. The goal was to develop a client-side only web site that behaved as much like a real web site as possible so as to avoid being distracting.

The simple task assigned to the user was to order two books – one fiction and one non-fiction, one CD – classified as “Alternative Rock,” and one 2004 release DVD. A user without TAMS would have to drill down into the site to verify that their items meet these criteria; the user with TAMS was given hints via a short series of alerts. The users were warned that if they guessed on the category of an item, they might be wrong. Figure 14 depicts a screen captures of this fictional web site.

The authors were surprised to find that some of the users involved in these experiments had no experience purchasing items from the web or filling out web-based forms while others had no trouble figuring out how to complete the assignment without having anything explained to them. For this reason, these experiments involving this task were run only with users that told us they were familiar with these tasks so that our results were not unduly biased by experience level.

3.4.1.3 Situational versus Peripheral Awareness Configuration

The Situational versus Peripheral Awareness Configuration was defined to help assess the Peripheral Awareness Enhancement MOE and the HMAS Interference with Situational Awareness MOE. The task involved having the user watch a sequence of four simulated UAV flight loops over Norfolk Harbor using the 3D display portion of CERRTS. Because it was desired that the users not become overly familiar with the Baghdad environment and thus biasing the results of future CERRTS experiments, CERRTS displayed 3D data of the Norfolk harbor region instead. One simple vignette was developed that described an emergency response to an incident involving a team of snipers randomly shooting civilians in the Norfolk harbor region. The user was asked to say “Norfolk” whenever they saw the word “Norfolk” appears onto the 3D display in front of them. The user was also asked to say “Sniper” whenever they saw or heard the word Sniper in an alert. This task enabled an assessment of how well the user was able to pay attention to the situation in front of them and the alerts in their peripheral vision at the same time. It also gave the users the ability to do a self assessment which clearly helped them express an informed opinion in responding to the user surveys.

Figure 14: Screen capture of fictional web site used for Simple Configuration.

In order to perform semi-automated data collection and analysis in a reliable and accurate fashion, a simple "Norfolk" GUI, depicted in Figure 15, was developed. Every time the user said the word "Norfolk," the person conducting the experiment would click the Norfolk button on this GUI, and every time the user said "Sniper" this person would click the Sniper button. Each button click triggered the writing of the event and its time (e.g. Norfolk or Sniper) into a log file. The Start and Stop button were used to insert markers in the log file to denote the Start and the Stop of each UAV flight loop. This enabled an accurate recording of the time at which the user called out these two words.

The data collection capabilities built into the alert generation module was used to collect ground truth on the actual time all "Sniper" alerts were transmitted. Data on the timing and quantity of the word "Norfolk" in the UAV flight loop simulations was manually collected and recorded.

Figure 16 depicts one of the alerts developed for the "Norfolk" vignette.

Figure 15: Simple Norfolk GUI which facilitated Data Collection and Analysis.

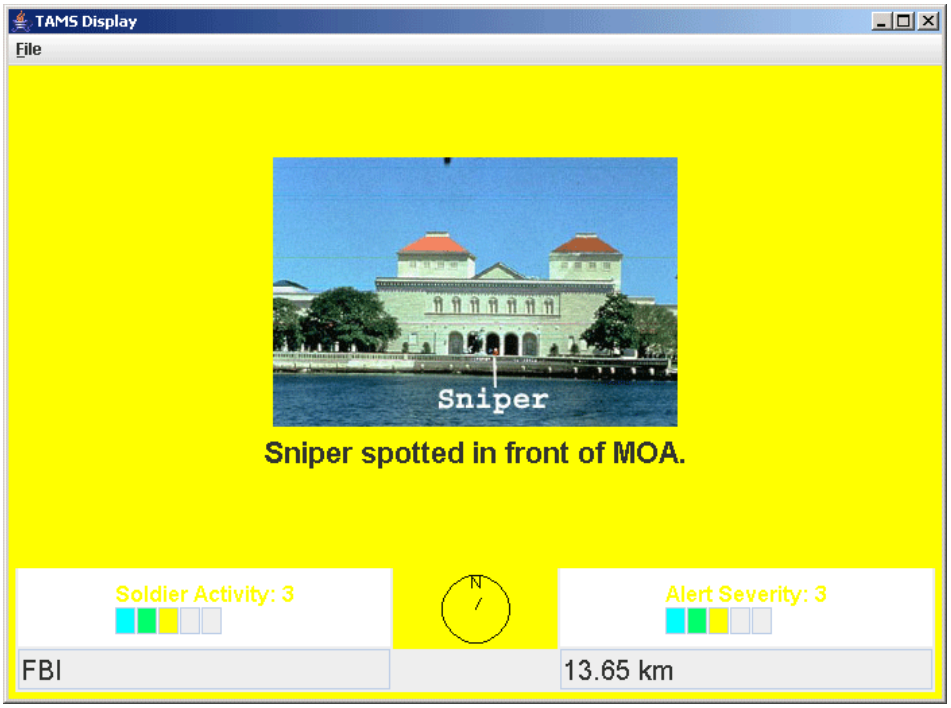


Figure 16: An example alert from the Norfolk vignette**.**

Afterward the user was presented with eight pictures and asked which ones they had seen in an alert, which gave another measure of their ability to understand their situation.

The vignette developed involved several emergency responder organizations responding to the sniper incident. A set of 108 alerts was developed that involved text, audio and imagery alerts. Raytheon was unable to find any video that was felt to be realistic in this vignette but it was not necessarily important to have video alerts. The IME was not used in this configuration to filter alerts, instead the alerts were allowed to stream directly to the alert presentation module. A new alert was presented every 5-10 seconds, depending on the content of the alert.

Every attempt was made to make the vignette as realistic as possible to avoid distracting the user. It was also felt that the more realistic the vignette, the more the user implicitly understood the purpose of TAMS, and the more fairly they were able to respond to user survey questions.

This experiment yielded very accurate results since it did not depend at all on a user's ability to perform a task for which they may or may not be proficient. It was certainly noted by observers of these experiments that the users respond differently to different alert presentation modes (i.e. text-only, normal, multi-modal).

3.4.1.4 Video Game Configuration

This task was meant to objectively support data collection against the MOP's associated with the Situation Understanding MOE. By developing an experiment configuration based on a novice-level off-the-shelf OIF-inspired urban combat video game[5], it was possible to immerse the user in a surprisingly realistic urban combat situation and collect data on the number of blue versus red kills, time required to complete a mission, etc. This task was very repeatable and the numbers were very objective.

To develop this mission, Raytheon ran through the video game several times, stored four different mission start points (a feature supported by the video game) and then developed four sets of alerts for each start point. The alerts contained text, audio, and images and warned the user where combatants, contraband, and other weapons were hidden, and recommended safer faster routes by which to reach the objective.

In order to prepare the user for the mission and insure that preparation was consistent across the user community, a "Smart Book" was put together that had screen captures of the objective and instructions on the mission. Figure 17 shows a page from the Smart Book for the mission with the "Neutralize" start point. The users were also allowed to use the built in training utility to learn to play the video game. Finally, the users were allowed to spend several minutes playing the video game before the experiments were started so that they could familiarize themselves with the operation of the controls.

---

[5] *Conflict: Desert Storm II – Back to Baghdad* by Gotham Games. See http://www.take2games.com/index.php?p=games&title=cds2

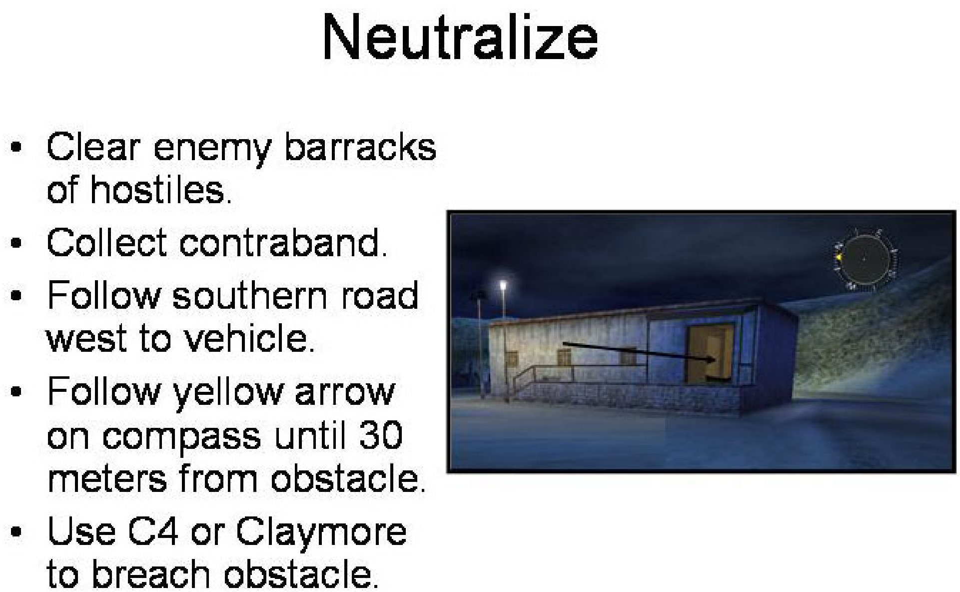


Figure 17: A page from our Smart Book that instructs the user on how to carry out the mission with the "Neutralize" start point.

It was found that the video game was much harder for some users to play than others, based on prior experience. Experiments based on the Video Game Configuration were run several times, with and without TAMS, and it was found that most novice users could not successfully complete the mission at all without TAMS, but were able to complete it with TAMS. However, since the video game does not display the statistics for a mission that was not successfully completed, all statistics for users who could not successfully complete the mission were discarded. The aim here is to identify a more consistent set of users upon which to base our statistics. In order to address some missed metrics, three users were added, who were experienced video game players, to the schedule. These users only ran experiments based on the Video Game Configuration.

Since there was no programmatic interface to the video game, the person conducting the experiment was required to synchronize the alerts with the video game manually, using the pause/play button on the alert generation module.

3.4.1.5 Hostage-Rescue Configuration

This task was primarily used to support data collection against the MOP's associated with the TAMS Training MOE. Two hostage rescue vignettes were developed for this configuration: one with two small unplanned events and one with one small and one big unplanned event. Similar to all current military training simulations that Raytheon is aware of, CERRTS does not fully support highly interactive vignettes that require a user to enter buildings. For this reason, this configuration required a "partner" to work with the user and inform them of certain events that CERRTS was unable to provide visualizations for on the 3D virtual display. The user was able to move from place to place, including going into buildings, but was unable to shoot anyone. The

partner assessed whether or not the user was able to move through the virtual world and follow their mission. To prepare them for their mission a “Smart Book” was put together for the Hostage Rescue mission and explained to them prior to the experiment. Users were also allowed to work with CERRTS to get a feel for moving around in the 3D display. Figure 18 depicts a page from this smart book.

Upon running experiments based on this configuration, Raytheon opted not to time the mission as originally planned, because it was found that some users were having a lot more trouble moving around the 3D display, based on previous experience or lack thereof with similar systems. Furthermore it was found that users with actual combat experience took much longer to complete their missions than users without because users with combat experience were much more careful in how they moved through the virtual space in that they avoided traversing tightly constricted spaces and approaching doors head on.

3.4.2 Experiment Schedule

Prior to the start of the experimentation campaign, a schedule of experiments was developed that involved bringing the same two users in for short periods of time at several intervals during the three day campaign. During the first day, a decision was made to alter the schedule because it was felt that the complicated sequencing of users through the experiments was far too risky since such a sequencing effort would have required an extreme amount of coordination to get the right users to participate at the right time.

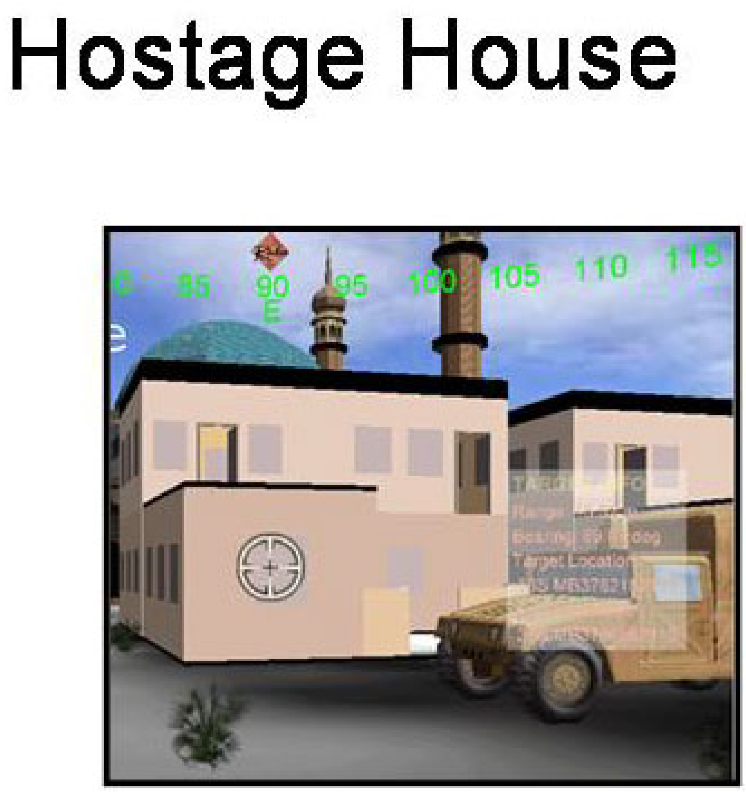


Figure 18: A page from the Smart Book that helped to explain the mission to the user.

Instead, the experimentation campaign was divided into five time slots (Monday PM, Tuesday AM, Tuesday PM, Wednesday AM, and Wednesday PM) with each user being assigned to a single time slot. A sequence of experiments was run with two users during each time slot. A few video game users were added to the schedule Tuesday after the regularly scheduled experiments were over and Wednesday after the regularly schedule experiments were over.

3.5 Experiments

The objectives of this contract required Raytheon and HRL to conduct a multi-day experiment and demonstration to which DARPA and AFRL were invited. In this section a detailed discussion of the methods and results achieved is presented. Raytheon integrated the TAMS prototype modules and the Experiment Configurations in a suite of labs in their Springfield, VA facility during the week preceding the actual experiments. HRL helped Raytheon conduct the actual experiments.

Twelve users were enlisted to support the experiments as test subjects. The users were all Raytheon employees working in the Springfield, VA facility. Over half had prior military experience, and about half of those had general infantry or equivalent training. Over half have spent time working with U.S. Central Command (CENTCOM), the Coalition Provision Authority (CPA), or Multinational Forces Iraq (MNF-I) in Iraq, either as a Raytheon employee or while still in uniform. Every user had spent time working one-on-one with warfighters. Their current role in the company was as System Engineers, System Administrators, Quality Assurance support, and a Chief Scientist.

The experiment schedule was organized in such a way as to be able to collect data against the MOP's that had previously been defined. Over the multiple day campaign, users were brought in two at a time and each user performed three to five individual experiments. Prior to starting the experiment, a high level discussion of TAMS was presented and the users were allowed to familiarize themselves with the hardware and user interfaces.

For each experiment, data was manually and electronically collected as mandated by the MOP's being evaluated by the specific experiment.

Overall, Raytheon and HRL felt that the three day experimentation campaign was a success. Some important quantitative results and some invaluable qualitative results were collected. The overwhelming enthusiasm for the technology expressed by almost all of the users was surprising. The TAMS prototype was extremely robust during the demonstration. The interfaces between the three software modules were robust enough to allow minor improvements to be made to each module over the course of the campaign.

The biggest issue experienced with the experiments was the complexity of them. Each new set of users had to be trained on the TAMS technology and the individual experiment configurations. Some manual data collection was required for several of the experiment configurations. Manual synchronization of alerts to the user's progress was required for three configurations. Log files

had to be appropriately archived after each set of users. Users required assistance with their hardware. Finally, there was constant pressure to keep to the schedule. This was an awful lot for the TAMS support team (two Raytheon and one HRL employee that supported the experiments) to cover. Plans for bringing in some engineers to participate as a surge team were abandoned when it was realized that it would probably take too much time to train the surge team on the tasks that needed to be performed and that there was a risk of newly trained individuals making a mistake and compromising an experiment.

The experiment schedule was anticipated to be aggressive so that, in addition to developing a Smart Book for the users, a Data Collection and Analysis (DCA) Smart Book was put together for the TAMS support team. This smart book contained checklists of the data that was supposed to be collected and the tasks that were supposed to be performed. Figure 19 illustrates a picture of a page from the DCA Smart Book. This greatly facilitated the robust performance of each experiment.

In the future, however, it is advisable that more time for the experiments is allowed and one full day of dry run experiments be scheduled to make sure that all of the kinks have been worked out of the system before commencing with the actual experiments.

The second biggest issue experienced was with the complexity of the tasks that the user was required to perform. It was not anticipated that the tasks associated with the Simple Configuration, the Hostage-Rescue Configuration, and the Video Game Configuration would be as difficult as they turned out to be for some of the users. This complexity did have some benefit however, it allowed for a more meaningful evaluation of peripheral awareness than might have been expected as the tasks really were requiring an extreme amount of focus from the users and the tasks really did allow the users to immerse themselves in realistic situations, especially in the case of the urban combat video game.

In the future, Raytheon recommends trying to develop the experiments around a military venue rather than trying to effectively develop a virtual reality from scratch. In particular, an appropriate venue might be a military operation on urban terrain (MOUT) area where the users can react with a physical environment rather than a computer keyboard and mouse. It is felt that this would allow a much more accurate assessment of the user’s operational performance and not their computer skills.

The military has stood up several MOUT areas in the U.S. including Ft. Irwin, Camp Lejeune, Quantico, Ft. Benning, Ft. Sill, Ft. Hood, Ft. Bragg, and Ft. Campbell. These areas generally involve mock cities that have actual buildings, streets, and cars. These areas are primarily used for MOUT training but they might be able to support experiments during breaks in the training schedule.

Metric	JB	PF	MC	DJ
Shots Fired				
Total Hits				
Headshots				
Torso				
Limbs				
Accuracy				
Kills				
Infantry				
Vehicles				
Times Hit				
Times MIA				
Mission Time				
Player:				
Time/Date:				
Vignette:	Neutralizer			
	Breakthrough			
	Clear Building			
	Kill the Tank			
TAMS Status	Filtered			
	Unfilterd			
	Text Only			
	Multimedia			
	None			
Video Game				

Figure 19: An example page from the DCA Smart Book that guided the recording of information associated with the video game task.

Ft. Sill also has the Joint Fires and Effects Trainer System (JFETS) which has been stood up to support training infantrymen on the Rules of Engagement (ROE) prior to deploying to Iraq. The JFETS consists of a mock Iraqi apartment from which the soldier can perform the duties of a sniper. The mock window looks out onto a large screen display on which is displayed an interactive virtual reality of an urban area. The room is also wired for surround sound to create realistic sounds of small arms fire and explosions. The current state of the technology is the user is unable to see into buildings or cars and the user is unable to affect objects or people in the virtual reality. Also, people in the virtual reality do not shoot at the soldier. Otherwise, the virtual reality is very good and in particular, the “crowd modeling” seems very realistic.

Under internal funding, Raytheon is currently evaluating the Gamebryo 3D graphics engine on which the JFETS virtual reality is based. This engine, which is available as a commercial product, has a C++ Application Program Interface (API) that allows a software developer to create custom applications. Raytheon is also evaluating the America’s Army computer game developed by the U.S. Army. This game is available for download and is customizable. The goal

of the evaluation is to assess how much effort is required to develop a customized virtual reality to support future urban operations experiments.

There are many tradeoffs involved in determining the best venue in which to conduct an experimentation campaign. The physical MOUT areas are probably the most realistic but they provide no inherent means to generate the alerts for TAMS to use. If a follow on phase addressed alert generation in addition to alert management, then the TAMS infrastructure might be used to address this gap.

Developing a game-like virtual reality might require more effort but experiments are potentially more automatable and predictable, it would probably take fewer people to execute such an experiment, and execution of the experiments would not likely be dependent on the training schedule of a MOUT area. Furthermore, if there is an API available, then it should be straightforward to build in automated data collection

Some minor issues during the experiment were experience because one of the TAMS laptop crashed during some experiments and as a result, the data collection log files were not archived. This issue did not become apparent until after the experiments concluded. In the future, the data collection software should be designed to write to disk continuously rather than buffering data and not writing until the buffer is full.

3.6 Analyses

The objectives of this contract required Raytheon and HRL to develop an AAR report that contains AAR results compiled for the TAMS prototype. This subsection presents this AAR report. The AAR activities involved entering the collected data into Excel spreadsheets, either automatically from log files (usually after some format editing was performed) or manually from DCA Smart Book pages and then using Excel formulas to process the data in such a way that it addressed the MOP's. In the following subsections, the AAR experiment report is presented, organized by MOE.

3.6.1 MOE 1: Peripheral Awareness Enhancement

This MOE was addressed with the Simple Configuration, the Video Game Configuration, the Hostage-Rescue Configuration, and a user survey. Users responding to the written survey based their comments on all of the experiments in which they were involved. The Video Game Configuration results are presented in Tables 6 – 9. This Video Game Configuration explicitly addresses three of the five MOP's associated with this MOE: "Amount of time saved by having TAMS", "Number of red kills with TAMS and without TAMS", and "Number of blue kills avoided with TAMS". Implicitly it addresses the MOP "% of situations soldier reacts appropriately when he encounters an unexpected event." It was observed that users with TAMS were much less likely to react inappropriately to an unexpected event, but no attempt was made to quantify this observation.

Table 6: Results of experiments from Breakthrough Mission for our Video Game Configuration

	No Alerts	Audio-Only Alerts	Multimodal Alerts
Shots Fired	168	95	184
Total Hits	26	35	54
Accuracy	15%	37%	29%
Kills	1	2	6
Infantry	1	2	6
Vehicles	0	0	0
Times Hit	105	110	107
Times MIA	1	0	1
Mission Time	9:06	8:11	5:31
Team Shots Fired	941	818	1024
Team Total Hits	179	128	193
Team Accuracy	19%	16%	19%
Team Kills	44	36	43
Team Infantry	44	36	43
Team Vehicles	0	0	0
Team Times Hit	328	310	346
Team Times MIA	1	0	2

The results of these experiments based on the Video Game Configuration are presented in precisely the same way that the video game presents the results on a mission (rather than in terms of the discussed MOP's). It was felt that this was appropriate because it leaves less room for interpretation.

The game was set up with a single person actually playing, but there were three other (virtual) squad members that were assigned to support the team player. Given that every user was playing this game for the first time, it had to be assumed that they were mostly focused on their own performance during the video game and not working strategically with their teammates (other than avoiding letting them get shot). For this reason, it was felt that many of the team statistics are somewhat unrepresentative.

While the results are somewhat mixed, in aggregate, they do show that using TAMS saves time, and avoids blue kills. In this video game, when a soldier goes down he is declared MIA. It is then up to the player to go save his teammate. If the player does not save him within a certain time limit, the mission fails and it is not possible to gain access to mission statistics. During the experiment, the users completed all missions when using TAMS. In a couple instances, they failed their missions without TAMS. In these cases, we ignored the experiment in the AAR.

Table 7: Results of experiments from Neutralize Mission for our Video Game Configuration

	No Alerts	Video-only Alerts	Multimodal Alerts
Shots Fired	82	176	175
Total Hits	9	44	32
Accuracy	11%	25%	18%
Kills	3	2	4
Infantry	3	2	4
Vehicles	0	0	0
Player Times Hit	79	81	32
Player Times MIA	0	0	0
Mission Time	8:38	7:24	5:27
Team Shots Fired	525	665	445
Team Total Hits	99	175	99
Team Accuracy	19%	26%	22%
Team Kills	32	63	24
Team Infantry	32	32	24
Team Vehicles	0	29	0
Team Times Hit	413	225	96
Team Times MIA	2	1	0

The Hostage-Rescue Configuration was used to assess the MOP: "% of situations soldier reacts appropriately when he encounters an unexpected event." With TAMS, the user reacted appropriately 92% of the time when he or she encountered an unexpected event. Without TAMS, the user reacted correctly 17% of the time when he or she encountered an unexpected event. In observing these experiments, the TAMS support team believed that the user without TAMS became completely immersed in the task of navigating through the CERRTS 3D virtual reality and were ignoring almost everything else. The user with TAMS was far more vigilant and was generally paying more attention to everything.

It was felt that the Hostage-Rescue Configuration was not ideal for assessing this MOP. Something along the lines of the JFETS system, discussed in the previous section, is probably more appropriate for this type of experiment because it involves more than just interaction with a single computer and therefore would force even the user without TAMS to be more aware of their surroundings. The TAMS support team was surprised to find that all of the users were having difficulty navigating through the 3D display, especially when trying to enter buildings and walk around rooms, and this posed a big distraction for everyone.

Table 8: Results of experiments from Kill the Tank Mission for our Video Game Configuration

	No Alerts	Audio-only Alerts	Multimodal Alerts
Shots Fired	713	695	639
Total Hits	162	161	139
Accuracy	23%	23%	22%
Kills	35	32	31
Infantry	34	31	30
Vehicles	1	1	2
Times Hit	352	377	383
Times MIA	2	1	2
Mission Time	27:38	25:26	25:08
Team Shots Fired	2097	1934	1927
Team Total Hits	475	508	506
Team Accuracy	23%	26%	26%
Team Kills	122	114	112
Team Infantry	120	113	114
Team Vehicles	1	1	1
Team Times Hit	1075	854	836
Team Times MIA	2	1	2

For the users with TAMS, the alert presentation mode for the Hostage-Rescue Configuration was varied, considering normal TAMS, multi-modal TAMS, text only TAMS, and unfiltered TAMS. For these experiments, it turned out not to make a difference for the MOP that was being assessed. It was noticed that it was taking the user with text-only TAMS longer and the users with unfiltered TAMS even longer to react to unexpected events, but the nature of the experiments were such that there was no real time penalty.

The Simple Configuration (ganges.com) was used to assess the MOP "Amount of time saved by having TAMS." On average, it took the user with TAMS 126 seconds to perform the task and the user without TAMS 217 seconds to perform the task. This is consistent with the results collected on the Video Game Configuration.

Data was not explicitly collected for the "Preferred alert layout" MOP. The reason for this is that it seemed more appropriate to work with the users and update the alert layout based on their suggestions directly rather than conducting an experiment to indirectly access preferences. There were some very good suggestions that were beyond the scope of what could be accomplished once the experimentation campaign had started. Table 5 reflects some of these comments.

The final MOP for this MOE is the results of the user survey. Table 10 summarizes these results.

Table 9: Results of experiments from Clear Building Mission for our Video Game Configuration

	No Alerts	Video-only Alerts	Multimodal Alerts
Shots Fired	187	355	231
Total Hits	15	101	29
Accuracy	8%	28%	13%
Kills	4	18	4
Infantry	4	18	4
Vehicles	0	0	0
Times Hit	48	151	27
Times MIA	1	0	0
Mission Time	18:38	13:14	7:01
Team Shots Fired	264	1306	522
Team Total Hits	56	283	125
Team Accuracy	21%	22%	24%
Team Kills	15	64	35
Team Infantry	15	64	35
Team Vehicles	0	0	0
Team Times Hit	113	467	85
Team Times MIA	3	0	0

Table 10: Results of User Survey for Peripheral Awareness MOE.

	Responses *				
Question	**Avg.**	**σ**	**Min**	**Max**	**# Users Responding**
TAMS will aid a soldier in being aware of things going on in his periphery while he is focused on completing a task or mission.	8.00	1.83	5.00	10.00	8 of 8
It is important for a soldier to be aware of things going on in his periphery (the activities of adjacent, supported, supporting units, etc.) while he is focused on completing an urban operation or urban combat task or mission.	8.56	2.18	5.00	10.00	8 of 8
TAMS represents an improvement over how a soldier would currently be made of things going on in his periphery.	7.86	2.45	3.00	10.00	6 of 8

* Scale for responses: 1 – 10, where 1=strongly disagree and 10=strongly agree

In observing the experiments the TAMS support team felt that the users' peripheral awareness was increased, perhaps more than they were aware, by TAMS. On some occasions users complained that they were not getting enough alerts and on a couple occasions it was noticed that the less experienced video game players were doing better than the more experienced video games players because they were depending more on their TAMS prototypes than the more experience players.

3.6.2 MOE 2: HMAS Interference with Situational Awareness

This MOE was primarily addressed with the Situational versus Peripheral Awareness Configuration. The MOP's for this MOE are as follows:

- % of time soldier properly deals with important information in front of him (not on goggles)

- % of time soldier properly deals with important information on periphery (only on goggles)

- Time required to react to alert on goggles.

- User Rating

In order to collect against the first three MOP's, several users performed the task associated with the Situational versus Peripheral Awareness Configuration twice: once in either normal or text-only alert mode and the second time in multi-modal mode. Unfortunately, due to the problem experienced with one of the TAMS laptops crashing and corrupting log files, no complete log files from the Text-only experiments were preserved.

Having observed the users run the text-only alert modes, it was clear that they were struggling to switch focus between reading alerts on the goggles and watching the simulated UAV flights through Norfolk on the large computer display. Unfortunately, it is not possible to present metrics to back up this observation.

Table 11 below shows the tabulation of the results collected against the Situation versus Peripheral Awareness configuration experiments to address the first three MOE's.

The results for the category **Missed Norfolk** appear quite bad. The reason for this was primarily because this was a very hard task to perform. The user was asked to call out the word "Norfolk" every time they noticed the word "Norfolk" appears on the screen as the virtual UAV simulated a flight over Norfolk harbor.

Table 11: Tabulation of results for Situation versus Peripheral Awareness configuration

	Normal	**Multimedia**
Average Reaction Time	3.33 sec.	3.83 sec.
Missed Sniper	5%	2%
Missed Norfolk	58%	48%
False Sniper	3%	1%
False Norfolk	3%	3%

While no formal plans were made to test users on this configuration with no TAMS, informal tests yielded surprisingly bad results. As the simulated UAV flew a loop over the harbor, the word could scroll onto either the left side of the screen or the right. It was simply very hard to watch both sides simultaneously.

The rest of the results are very consistent with what was observed. The user generally did better when the alerts were presented in multi-modal mode, although it did take longer for the user to react. The observed reason for this is that the user was waiting for text alerts to be read to them by the text-to-speech synthesizer rather than read the text alerts themselves.

While it was initially thought that the user might find the multi-modal alerts distracting and hence do less well on the "% of time soldier properly deals with important information in front of him" MOP, it was observed that the user actually did better.

The TAMS support team felt that the reason for this was that the user was able to better focus on the 3D display in front of them when they received an audio cue for each new alert. Hence the user did not have to keep looking back and forth between the 3D display and the goggles to see if a new alert had been presented.

Table 12 provides a summary of the user responses to the survey associated with this MOE.

This survey was given after each user completed all of his or her experiments so these responses reflect his or her experiences on multiple experimental configurations. While most of the users found TAMS to be somewhat distracting, the results from other surveys and written comments reflect that they accept this so long as the alerts are well filtered, especially when the user is extremely busy.

Table 12: Results of User Survey for HMAS Interference with Situational Awareness MOE.

	Responses *				
Question	**Avg.**	**σ**	**Min**	**Max**	**# Users Responding**
TAMS DID NOT distract me while I was Conducting the experiment.	6.33	2.91	3.00	10.00	9 of 9
I was able to focus on my mission and ignore TAMS when the situation before me required my full attention.	8.33	2.00	4.00	10.00	9 of 9
The TAMS see-around display is effective.	7.33	1.95	3.00	10.00	8 of 9
I would want the ability to quickly move the display away from my eye in an actual Urban Operation or Urban Combat situation.	7.44	3.40	1.00	10.00	9 of 9

* Scale for responses: 1 – 10, where 1=strongly disagree and 10=strongly agree

3.6.3 MOE 3: TAMS Training

This MOE was primarily addressed with the Training Configuration. The following MOP's were originally developed to assess it:

- Convergence rate of IME
- Learning rate (also referred to as Training rate)
- Accuracy
- % "drift" in IME rules over time
- % improvement of high-activity IME training over low-activity IME training?
- Compare different training methods
- User rating

For the final campaign, not all of these were assessed as originally planned. Learning rate and Accuracy were accessed using an automated experiment (no users involved). The "% drift in IME rules over time" was not assessed due to schedule constraints. The next two were combined by letting the user train their instantiation of the IME in different was no appreciable difference was seen.

To understand how learning rate and accuracy were measured, these terms are defined as follows:

- *Accuracy:* Accuracy is defined as the number of alerts correctly presented divided by the total number of alerts. A value of 1 means all of the alerts received by the system were presented correctly. A value of 0 means the IME presented no alerts correctly. The IME

deciding not to present an alert is considered a correct action, so if a soldier wanted to turn off all alerts and the IME learned not to show any, the accuracy would be 1.

- *Learning rate:* The IME goal is to have high accuracy with a minimum of user input required. A learning rate of 1 is best, while a learning rate of 0 means learning is too slow for the level of accuracy achieved. So learning rate is defined as the accuracy minus the quotient of the number of user inputs and the possible number of inputs.
  Learning Rate = Accuracy - (number user inputs / possible number inputs)

The Learning rate is capped at 0 so there are no negative values. For example, if no questions are (i.e. use case histories) asked and there is an accuracy of 1, then the learning rate is highest at 1. If all possible questions are asked and there is an accuracy of 1, the learning rate is 0, meaning it took too much input. If no questions are asked and none are correct, the learning rate is 0. If 10% of possible questions are asked and 90% of them are correct, the learning rate is 0.8.

For the Convergence Rate MOP, the reduction in incorrect choices was measured. This metric only provides an indication that the IME is learning the user's preference, and is only useful in the context of a vignette or possibly in an operational setting. This type of training session has the value that it can more accurately represent the context in which an alert is received. On the other hand, there is no control over the specific order of the alerts since they flow based on the simulation (or live operations).

The automated experiments for measuring learning rate and accuracy required a truth set and were either ontology-based or rule-based since the characteristics of the alerts and interface modes were predefined. A truth set is as a set of alerts for which the "correct" IME responses have been predetermined by observation. In the case of a rule-base, a tool to take descriptive rules from a soldier and generate a truth set from the rules was used. This method has the significant disadvantage that the soldier must be able to define preferences across a five-dimensional input space, and therefore is not a good method for constructing an actual decision classifier. However, it is useful for assessing the IME's ability to learn a pre-defined truth set.

For ontology-based methods, "presentation graph" reasoning was applied to the ontology to decide which training examples are the best to ask about to maximize the learning rate. This approach requires active learning methods that were not addressed in this program but should eventually be considered.

In experiments that were conducted using the truth set training architecture, it was found that the percentage of correct choices for display was improved by the choice of which initialization was used, independent of the number of training samples. Table 13a defines the truth set training rules for alert severity and Table 13b defines the truth set training rules for soldier activity. It was also found that as the number of training samples increased, the accuracy of the system improved (e.g. from 69% for only 10 training samples to 94% for 150, shown in Tables 14a and 14b). In these tables, results of two experiments are shown, each with three examples based on different initializations: always display, never display, and display under conditions defined by a set of rules.

Table 13a: Truth table set 1: emphasis on alert severity

Alert Severity and Condition	Required Action
If Alert Severity is 5,	produce alert
If Alert Severity is 4, unit generating alert is 2 or better and Activity Level is 4 or lower,	produce alert
If Alert Severity is 3, unit generating alert is 3 or better, Activity Level is 3 or lower and Mode is unpunished,	produce alert
If Alert Severity is 2, unit generating alert is 4 or better, Activity Level is 2 or lower, Alert Distance is Near and Mode is unpunished,	produce alert
If Alert Severity is 1, unit generating alert is 5, Activity Level is 2 or lower, Alert Distance is Near and Mode is unpunished,	produce alert

Table 13b: Truth table set 2: emphasis on soldier activity

Alert Severity and Condition	Required Action
If all of severity, unit, and distance are greater or equal to (activity – 1) and some of them are greater than or equal to activity level,	accept all interface modes.
If activity is 5,	do not accept video mode.

Table 14a: Experimental results for initialization bias from historical reference models emphasizing alert severity

# of samples Initial state	10	50	100	150	200	250	300
All accepted	1740(69.6%)	2213(88.5%)	2332(93.3%)	2373(94.9%)	2345(93.8%)	2368(94.7%)	2358(94.3%)
Truth table	1537(61.5%)	1865(74.6%)	2119(84.8%)	2198(87.9%)	2219(88.8%)	2229(89.2%)	2248(89.9%)
All rejected	1850(74.0%)	1845(73.8%)	2003(80.1%)	1901(76.0%)	2021(80.8%)	2022(80.9%)	2045(81.8%)

Table 14b: Experimental results for initialization bias from historical reference models emphasizing soldier activity

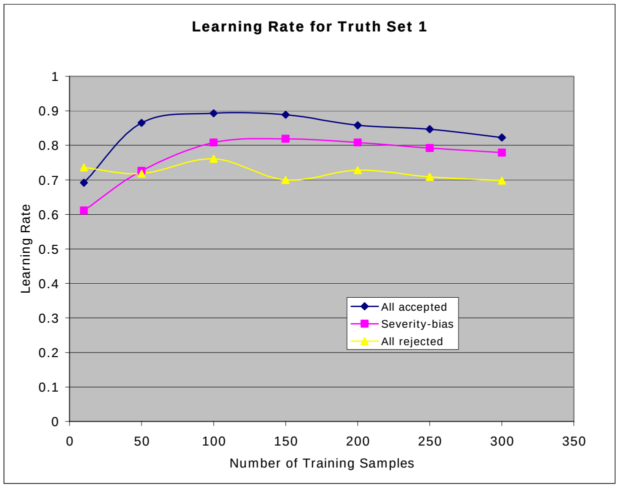
# of samples Initial state	10	50	100	150	200	250	300
All accepted	1859(74.4%)	2308(92.3%)	2370(94.8%)	2405(96.2%)	2400(96.0%)	2393(95.7%)	2417(96.7%)
Truth table	2076(83.0%)	2324(93.0%)	2339(93.6%)	2356(94.2%)	2363(94.5%)	2346(93.8%)	2361(94.4%)
All rejected	1962(78.5%)	2191(87.6%)	2158(86.3%)	2091(83.6%)	2117(84.7%)	2095(83.8%)	2113(84.5%)

Figure 20 shows the graph of the learning rate for Table 14a and 14b. Note that the learning rate improves rapidly up to a point and then levels out or drops.

Figure 21 details the results of the Training Configuration experiments. In these experiments, one complete training vignette was run, the IME was retrained, and a second complete training vignette was run. Figure 21 plots user corrections from the first vignette to the second vignette.

In aggregation, the users averaged 11.43 corrections for the first training vignette and 10.43 corrections for the second training vignette. The IME training algorithm is converging, as desired.

On an individual basis, however, about 33 % of the users made no corrections during the second training session, 33 % made fewer corrections in their second training session than their first, and 33 % made substantially **more** corrections in their second training session than their first. It is felt that this points to a user training issue and not a technology issue. It has been hypothesized that these users did not really understand the training procedure and how to use it effectively until the second training run.



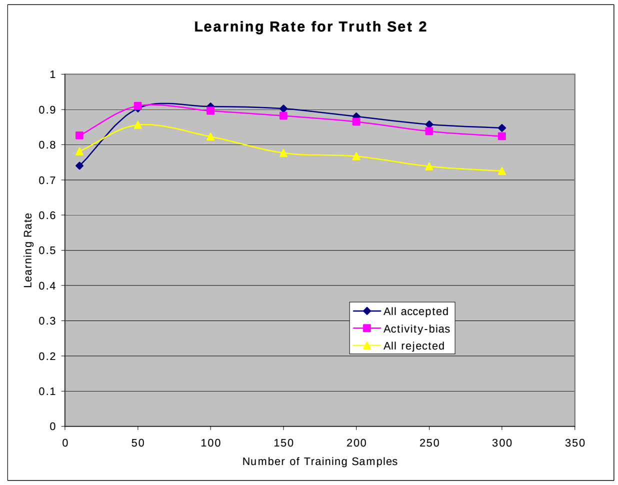


Figure 20: The learning rate for truth sets defined in Tables 13a and 13b based on a total possible training sample size of 2500 alerts.

Figure 21: User corrections versus time for two supervised batch training runs**.**

This situation might be improved by allowing more time for user preparation during the experiments and by implementing an interactive training algorithm (that reacts to each training input) so that the user implicitly sees the effect of their previous corrections.

The training rate (how long it took for a user to respond to an alert) averaged 7.164 s on average for the first training vignette and 7.160 s for the second training vignette, with about 33% of the users taking longer to respond to an alert on the second training vignette, 33% of the users taking less time to respond to an alert on the second training vignette, and 33% of the users taking about the same amount of time.

Table 15 tabulates the results of the user survey for this MOE. The minimum numbers probably reflect the users who were not as well prepared for the training experiments by the TAMS support team.

3.6.4 MOE 4: Situation Understanding

This MOE was primarily addressed with the Situation versus Peripheral Awareness configuration (also referred to as UAV flight or Norfolk) experiment configuration. The MOP's originally defined for this MOE are as follows:

- % of retention of alerts for different presentation modes.
- % improvement in understanding situation
- User rating

Table 15: Results of user survey on training.

	Responses *				
Question	**Avg.**	**σ**	**Min**	**Max**	**# Users Responding**
Each soldier should be allowed to "train" his own TAMS.	8.00	3.32	5.00	10.00	8 of 9
The TAMS training process was easy to understand.	8.89	0.74	8.00	10.00	9 of 9
On a scale of 1-10, how important is ease of use for introducing enabling technology to soldiers?	9.67	0.47	9.00	10.00	9 of 9
I **DID NOT** find the TAMS training process annoying.	8.44	2.27	3.00	10.00	9 of 9

* Scale for responses: 1 – 10, where 1=strongly disagree and 10=strongly agree

The Situation versus Peripheral Awareness configuration was used to assess the first MOP because there were many more alerts with this configuration than with the Hostage-Rescue Configuration. For this task each user was shown screen captures of eight image alerts, three of which he or she had seen during the vignette and was asked to identify those image alerts he or she saw. 75% correctly identified all three, the remaining 25% correctly identified two out of three. No one incorrectly identified an alert they did not see.

For the second MOE each user was shown four images from the hostage rescue vignette and asked to specify the significance. The users without TAMS correctly identified the significance of two of the four images and the users with TAMS, regardless of mode, correctly identified all four. Table 16 tabulates the results of the user survey for this MOE. The responses to the last question were fairly diverse. We feel that the question was too vague to elicit a consistent response.

3.6.5 Miscellaneous Results

Overall the TAMS support team felt that the results were fair, objective, and mostly consistent with observations and expectations. Perhaps the biggest surprise is the popularity of multi-modal alerts over all other alert types and the difficulty in paying attention to text alerts. Table 17 presents some miscellaneous comments extracted from the user surveys.

Table 16: Results of user survey on situation understanding.

	Responses *				
Question	**Avg.**	**σ**	**Min**	**Max**	**# Users Responding**
TAMS will improve a soldier's overall understanding of his situation.	8.25	1.56	5.00	10.00	8 of 8
TAMS DID NOT confuse me about my situation.	7.63	2.50	3.00	10.00	8 of 8
More TAMS alerts are better than fewer TAMS alerts.	5.50	2.24	2.00	8.00	8 of 8

* Scale for responses: 1 – 10, where 1=strongly disagree and 10=strongly agree

Table 17: Miscellaneous comments from user survey

Voice recognition of Accept, Reject, and Don't Care.
Caution user on reliability of data as the users will become very dependent on it.
The ability for the system to remember the soldier's choice or preference of alerts is a great asset.
Build into helmet/visor system.
Receiving related alerts on demand (would make TAMS better)
System only worthwhile for video and image based alerts.
Map based visual cues to accompany image alerts would be very helpful.
Give user ability to "pause" alert or rewind.
Once trained to use this would be a very useful tool

Table 18 presents a summarization of the user survey results when only users with military experience were considered. It shows that the highest two scores were for the two MOE's that Raytheon and HRL feel are most relevant to the work conducted for this program: Peripheral Awareness Enhancement and TAMS training.

Table 18: Summarization of user survey scores for users with military experience only.

Measure of Effectiveness	Mean	σ	Min	Max
Peripheral Awareness Enhancement	8.43	2.70	6.67	10.00
HMAS Interference with Situation Awareness	7.17	4.64	3.50	9.33
TAMS Training	9.13	2.49	7.00	10.00
Situation Understanding	8.13	3.77	5.00	10.00

4. DISCUSSION

The detailed specific objectives of this program were to establish a ConOps for the HMAS and the associated experiments, to design and develop the IME, to design and develop a prototype TAMS, to design, develop, integrate and conduct experiments and finally to conduct AAR activities. The following subsections provide details of the accomplishments of this effort towards these objectives.

4.1 ConOps Investigation

The major accomplishment derived from the successful completion of this objective was the ability to carry out the Experiment Design and Development task as planned.

4.2 Information Management Engine

This objective, here entitled Information Management Engine, is a restatement of the contract Statement of Work (SoW) subsection "3.3 Task 3: IME Correlation and Aggregation Engine." While the objective of the IME task was accomplished, it was not done as originally planned. The originally proposed IME learning algorithm was based on one developed for the Intrusion Detection domain. Early on in the program it was determined that the Intrusion Detection based learning algorithm was not the most appropriate reuse candidate, so the IME was developed using a different approach. The major accomplishment derived from the successful completion of this objective was the ability to use the IME software prototype as the central component of the TAMS software prototype.

4.3 Tactical Alert Management System

This objective, here entitled Tactical Alert Management System, is a restatement of the two contract Statement of Work (SoW) subsections "3.3 Task 4: IME Display Engine and Task 5: IME User Interaction Engine." has been restated from two tasks that appeared in the original proposal: IME Display Engine and IME User Interaction Engine. While the ultimate objective of this task was accomplished, it was not done as originally planned. Feedback and actions taken from AFRL and DARPA during the kickoff meeting made it apparent that the original software reuse plan for these original tasks was not be suitable. Furthermore, the original plan was for HRL to perform these tasks. To help lessen the impact of the loss of a reuse candidate, Raytheon performed the implementation of what was originally referred to as the display engine, although HRL did contribute to its development and implementation.

Another change to this task not originally planned on at contract award was the need for the alert generation module (AGM). Original plans were for the majority of alerts being generated by the CERRTS system, but as the ConOps evolved, it was realized that this assumption was simply not consistent. Furthermore, as the experiment design evolved, it was recognized that a more rigorous, repeatable, and fully automated mechanism for generating all alerts was required.

The major accomplishment of this task was the creation of a fully integrated TAMS prototype that was used to support the experimentation campaign. The prototype was extremely stable and the interfaces between modules were defined well enough that we were able to make minor modifications to the various modules while the experimentation campaign was going on in order to immediately incorporate and evaluate user suggestions.

4.4 Experiment Design and Development

This objective was met but the plan was changed in reaction to direction received from DARPA and AFRL at the Integration Event. The original plan called a modification to CERRTS to send the alert format required by TAMS, which was done prior to the Integration Event. During the Integration Event, however, DARPA gave direction not to use TAMS to display C2 messages. All of the alerts sent by CERRTS are C2 messages with the exception of the “Plain Text” alerts.

All of the alerts used in the experiments can be sent as Plain Text alerts via CERRTS, however, it cannot do so in a repeatable and controllable fashion. Furthermore, the CERRTS interface that allows this to be done is cumbersome and very manual, so it would have been extremely difficult to user CERRTS to pass alerts in the quantities that were needed to support the experiments.

For these reasons, Raytheon opted to bypass the CERRTS alerting capabilities and expand a tool originally developed as a test tool to facilitate the integration of CERRTS to the IME. In so doing, not only was it possible to implement a repeatable, completely controllable mechanism for generating and sending alerts, it was also possible to develop a simple mechanism for manually creating, reviewing, and managing alerts.

Raytheon had originally planned on supporting two experiment configurations at contract award, a Training Configuration and a Hostage-Rescue Configuration. However, as the ConOps and MOE’s were developed, it was felt that it was necessary to introduce three additional experiment configurations to support repeatable, objective, and realistic data collection against some of the MOP’s.

The major accomplishment derived from the successful completion of this objective was the development of five complete experiment configurations which enabled Raytheon and HRL to thoroughly evaluate the TAMS technology. In addition, six complete urban combat vignettes and associated alert sets for three of the configurations were created. Finally, the six additional alert sets for the remaining two experiment configurations were created.

4.5 Integrate and Conduct Experiments

This objective was performed as originally planned at contract award. The three software modules that made up TAMS with the goggle mounted displays and a pair of headphones were integrated. Furthermore, the experiment configurations with TAMS were integrated. Experiments were conducted and data was analyzed.

The major accomplishment derived from the successful completion of this objective was the collection of qualitative and quantitative data that allowed us to assess the TAMS prototype.

4.6 AAR Activities

This objective was performed as originally planned at contract award. The major accomplishment derived from the successful completion of this objective was the analyses included in Section 3 of this document.

5. CONCLUSIONS

The goal of this program was to develop a TAMS prototype that relies on a proof-of-concept IME to filter the alerts that get presented to its user. Furthermore, the program was to perform an experimentation campaign and AAR to assess the design and military utility of TAMS.

TAMS provided value by greatly enhancing the users' peripheral awareness while they conducted simulated urban operations. The IME utilized a novel filtering algorithm that addressed the very common problem of introducing new technology to the operational user by exploiting a supervised learning mechanism to produce an extremely simple and intuitive procedure by which the soldier can configure their system. To simplify the supervised learning process, the IME combined other machine learning mechanisms with the supervised learning mechanism to yield a powerful learning algorithm that the users implicitly used to configure their system to suit their needs and preferences. The users performed this task without understanding the underlying technology or configuration.

The results of the experimentation campaign bear out that the TAMS technology has merit and deserves further exploration. In the course of a two and a half day experimentation, twelve users who had had no previous knowledge of the system were presented with the technology, instructed on the IME training process, allowed to train their own instantiation of the TAMS software, and asked to perform several experiments with and without the TAMS prototype they had trained. All of the users were able to appreciate the value of the technology and the underlying training process. They were enthusiastic about this technology and in the short span of a couple of hours actually came to depend on it to perform some of the tasks they were required to carry out. Raytheon and HRL believe this is an impressive demonstration for a prototype system and a somewhat contrived urban combat experimentation campaign.

Two issues greatly complicated the creation and execution of the TAMS experimentation campaign: the difficulty of creating a realistic, high fidelity urban combat simulation that allowed for a fair assessment of this technology, and the synthesis of meaningful, convincing example alerts. The first issue is well known and there are several efforts underway to address the lack of modeling systems that can adequately simulate urban combat situations. Promising efforts are currently exploiting computer gaming technology. Raytheon and HRL believe that the discussed approach, which involved utilizing several different configurations (including an urban combat video game) to experiment with different aspects of TAMS, was the best that could be done under the constraints of the contract and that it allowed for a meaningful assessment of quantitative and qualitative results.

The second issue is complicated by the fact that the most meaningful alerts for the TAMS prototype are image alerts and video alerts and it is difficult to synthesize realistic, high fidelity images and video clips relating to a chosen urban combat vignette. The use of synthetic imagery from the simulation system built to model the PSDS2 video input streams was investigated. It was found that this simulation system, which is focused on creating synthetic imagery for a collection of sensors based on the PSDS2 sensor characteristics, did not produce images that did not seem useful to a dismounted soldier. This is primarily because the synthetic images of

Baghdad are created from a relatively low-fidelity three-dimensional model of Baghdad. Furthermore the three-dimensional model was only of buildings, roads, and terrain features in Baghdad. The system used a simplistic crowd-modeling algorithm to add people into the imagery. However, they appeared almost as silhouettes – there are minimal facial and wardrobe features. Also, the crowd-modeling algorithm tended to placed individuals in unnatural positions (it almost looked like they are mulling about in formation).

To produce some meaningful image alerts, screen captures from the simulations used in the experiments were used. In some cases, other images were overlaid onto the screen captures – for instance to show the ODA team landing on the roof of the rendezvous point – or added textual annotations to the screen captures. Several images were also downloaded from the web from news reports, military training exercises, blogs, and other miscellaneous sources. Several clips of video taken either from Baghdad or MOUT areas were also downloaded and edited into a set of very short video streams that could be worked into the storyline of the vignettes. After inserting the image and video alerts into each vignette, gaps were filled by creating either text or audio alerts so that the users were actually presented with a story which helped them understand and react to the underlying vignette.

Raytheon and HRL believe that the result was good enough to get the value across to the users of the TAMS prototype. A few did comment, properly so, that there were too many text alerts and they felt more image and video alerts were necessary. This is an artifact of the experimentation campaign and not of the TAMS technology.

One aspect of TAMS that Raytheon and HRL sought to investigate was the presentation of alerts to the user. Along these lines, the IME would take an alert as input and translate its presentation mode to one of a few possibilities: the original input mode, an alternative mode, or none (meaning that it was not presented to the user). Independent of the IME, a configuration parameter was built into the TAMS alert presentation module that allowed the TAMS support team to specify normal alert presentation, text-only presentation, or multi-modal, which performed speech synthesis on all text (amplification fields in the case of image and video alerts) and sounded a short notification sound at the beginning of every alert. Simply by emitting either earphones or goggles it was possible to simulate audio-only TAMS (with all text being converted to synthesized speech) and video-only TAMS. With these possibilities, the TAMS support team was able to experiment with several different TAMS configurations, both formally and informally.

Experiments with text-only, audio-only, and video-only alerts always yielded worse results than normal TAMS alerts (text, audio, image or video) which were always worse than multi-modal alerts (text, audio, image, or video + augmentation such as text to speech). Furthermore, when the IME translated an image or video alert to a text alert, the user was annoyed. These experiments do not conclusively suggest that only multi-modal alerts should be presented to the user, but rather the presentation methods need to be further studied. During the Integration Event, several other presentation considerations were discussed for the longer term ConOps of TAMS – including flashing text, vibrating alarms, etc. Follow on efforts should look at the issue

of presentation methods that take into consideration a wider range of possible presentation configurations, bandwidth requirements, and covertness requirements.

During the Integration Event, a Raytheon military subject matter expert (SME) claimed that the soldier would always want to flip the display away from the eye when in the move-and-shoot situation. The urban combat video game used was primarily based on a move-and-shoot scenario. When the statement was posed to the user in a user survey “I would want the ability to quickly move the display away from my eye in an actual Urban Operation or Urban Combat situation” in a user survey the results were mixed, with most users either strongly agreeing and strongly disagreeing. The users that strongly disagreed actually did better at the video game task than the users who strongly agreed.

As suspenseful as the urban combat video game was, Raytheon and HRL do not want to suggest that it is in any way equivalent to making contact with a hostile insurgent in a real-life urban combat situation. However, we do want to suggest that a user’s comfort level may be more closely related to the general comfort level with computer technology and furthermore, that future warfighters may be more comfortable with computer technology than past warfighters. Also, we believe that see-through and see-around displays, in general, may not elicit this response as strongly as the occluding displays that many users think of when asked this question.

In summary, the experimentation campaign showed the TAMS prototype to be stable, robust, flexible, responsive, and well-received by the users. They made several astute observations based on past experience and the experience they had with the TAMS system. There was strong concurrence among the users of the motivating premises for applying the IME to the TAMS prototype -- TAMS needs to be extremely simple for a soldier to use if it is to be successfully transitioned into an operational environment.

6. RECOMMENDATIONS

In the following, the detailed recommendations presented will be divided into four categories based on the key efforts on this contract. Overall recommendations will first be made for follow on work in the area of tactical alerting systems. The overarching recommendation is that an untethered full-duplex system that allows a user to generate alerts in addition to receiving alerts needs to be addressed. Secondly, recommendations will be made for follow on work in information management. The third topic, alert generation and alert presentation, is closely related to the first two, but has been called out separately to more clearly delineate the issues.

Finally, recommendations will be made for experimenting with, integrating, testing, and fielding a TAMS system. Our overarching experimentation recommendation on this topic is to work with the military to develop a more realistic urban combat simulation that does not unduly depend on the computer skills of the test subject.

6.1 Full-duplex Tactical Alert Management

As the ConOps for TAMS has evolved, it has become clear that the most meaningful alerts will often be generated by the same soldiers who are receiving TAMS alerts. The soldiers securing the perimeter of an objective location will have distinct viewpoints of the objective and can provide meaningful alerts to other TAMS users about any unanticipated activity in their sector.

We recommend that follow on work address using TAMS to generate alerts as well as receive alerts. Such an investigation may consider addressing several interesting technical challenges. A key technical challenge is developing an efficient robust and efficient mechanism for performing multimedia alert distribution over a mobile ad hoc network (MANET), preferably using a multicast protocol. Here we are assuming that the TAMS users will be wearing an untethered version of TAMS that communicates with a TAMS gateway via a wireless communication network based on an 802.11b, 802.11g or 802.16 standards. This assumes the gateway to be a vehicle or even a stationary operation center which provides an uplink to higher headquarters.

For this program, the TAMS hardware implementation was limited to a tethered, presentation only system. Follow on efforts should work with other programs or with other commercially available hardware to address, at least for a proof of concept, the remaining hardware requirements to realize the full ConOps: biofeedback sensors, imagery, video, and audio creation capabilities for generating alerts, user input devices, encryption, and wearable computing platforms in addition to the wireless communication requirements. To ultimately be successful, TAMS will require a low power lightweight hardware suite that integrates easily with the soldiers other gear.

The ConOps for alert creation, specifically how much of the process can and should be automated, is another area for consideration. If a soldier snaps an image of a hostile approaching the objective location from their quadrant, how would the image be annotated, or would it? Does it first get transmitted to higher headquarters to be reviewed and vetted, or is it handled completely by the gateway? Is there any enhancement that can be performed automatically that

adds value, for instance, using speech recognition to automatically fill in the alert amplification field of an audio alert?

Another technical wrinkle is determining where the IME resides in the TAMS architecture and how it interacts with the gateway. For the purposes of this contract it was assumed that the IME was running at the gateway thereby filtering out alerts before transmitting to the user. Assuming the IME is not updated frequently by the soldier, this makes the most efficient user of bandwidth. These assumptions need to be revalidated. Scalability of the IME is also an issue. For n TAMS users, do you need n instances of the IME running?

Related to the ConOps issues are the user interface issues. Several questions remain about how to create a user interface that is simple to use but powerful enough to not distract the user. The user needs to be able to create alerts and also ought to be able to request a replay or more information on an alert they have just received.

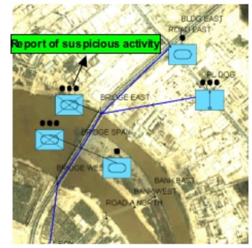
One possibility is to give the user a five button device, labeled “Back”, “Forward” (similar to a web browser) which automatically put TAMS into a review mode, “Pause” which will freeze the alert on the display and pause an audio or video clip, “Reset” which will take TAMS out of review mode and back to normal operations, and a “More” button. The “More” button might employ semantic technologies to rank all generated alerts with their relevance to the current alert (placing high priority over the most recently generated alerts) and then display the most relevant alert for the user. This acts as a very simple but potentially very powerful query interface to the TAMS alert stream. The act of pressing the “More” button on an alert might be fed back to the IME as a training parameter for the IME’s learning implementation.

Raytheon would expect to be able to exploit current and likely future work being done to ruggedize and extend Raytheon’s goggle mounted display product line. In particular, it is believed the work that is being done on the TUSK program and related programs might be used to form the basis for the TAMS hardware architecture. Current prototypes are addressing wearable processing and wireless communications. These capabilities could be extended to accommodate the needs of a wearable TAMS system.

6.2 Information Management

The implementation of the IME using a multi-class SVM with incremental updates enables the system to learn the user’s preferences fairly quickly for the types of alert input vectors that have been defined. In an operational system, the IME will need to function with a larger number of dimensions for both input and output. For example, the choice of output modes of only four types does not address many of the possible output options that a user may prefer under different circumstances. Table 19 lists a number of possible variations on output modes that may provide improved understanding of alerts and could be used in training the IME. Similarly, alert payload content should be reflected in the input dimensions. We recommend that further work address higher dimensional input and output vectors.

Table 19: Variations and examples of output modes the user may select during training.

Mode	Interface Options	Variables	Alert variations	Examples
Text	• No Display • Display Visual -Draw text at screen coordinates -In box -Across screen -Ticker (repeating) -Text icon at screen coordinates -Keyword at screen coordinates •Display Aural -Convert text to speech and play monaural -Convert text to speech and play spatial at world coordinates -Play "earcon" • Combination text display + audio	• Timing of display • Choice of location • Text size, style, color, (loudness) • Representation of icon (earcon) • Method of interaction with icon (earcon)	• Display relevant alert values as text • Severity maps to color (5=red, 1=green) • Unit displayed with icon • Show location with map or icon • Message newness maps to brightness of font (bold=new, dull=old)	S2 Report of suspicious activity Unit=2/6 Time=281100ZAPR2005 Location=38SMB38998580  Text only (top) and text on map (bottom)
Audio	• No Display • Display Visual - Convert to text and draw at screen coordinates - In box - Across screen - Ticker (repeating) - Keyword(s) • Aural -Play monaural -Play spatial at world coordinates -Play "earcon" • Combination audio + icon	• Timing of display • Choice of location • Audio loudness • Representation of earcon • Method of interaction with earcon	• Display relevant alert values as text (not good to use audio) • Severity maps to loudness • Unit displayed with icon • Show location with map or icon (best to use spatial) • Newness maps to ?	**Mode = Audio** Example = "Roof is clear"  Play audio (top) and link audio to earcon on map (bottom)
Image	• No Display • Display Visual - Draw at screen coordinates - In box - Relative to map (satellite image) - Relative to world -Draw icon at screen coordinates - Draw keyword at screen coordinates - Combinations (e.g. box, image, keywords, map)	• Timing of display • Choice of location • Image enhancements (e.g., reduced or enlarged) • Image overlays	• Display relevant alert values as text • Severity maps to color (5=red, 1=green) • Unit displayed with icon • Show location with map or icon • Message newness maps to brightness of font (bold=new, dull=old)	  Image in box (top) and combination of image with overlays (bottom)
Video	Same as image	• Timing of display • Choice of location • Playback speed • Video enhancements (e.g., reduced or enlarged) • Video overlays	• Embed values in video • Use other methods as before	 Sarnoff Video Flashlight Sarnoff Video Flashlight

Thus far, it has been assumed that the user's activity level is obtained through human observation. With additional input from biometric sensors such as body temperature, heart rate, respiration, movement, etc., the IME could learn a user's activity level in either a supervised mode during training or unsupervised during training or operations. It has also been assumed that the unit and location values are determined using a "policy" based method in advance, and then when the alert is received by the soldier, these values are computed relative to the soldier's unit or location (and are different for each soldier). The soldier may prefer to establish their own policy, which would enable them to prioritize alerts differently for display than the policy.

Future IME functionality should move toward an active learning strategy combined with historical reference models. Active learning may be described as a method for minimizing the size of the training set while maximizing the value of each training example. This is analogous to an excellent teacher who is able to choose the training examples that will be fastest for a student to learn while providing the most understanding of the subject. Active learning systems must maintain an accurate model of the learning system's state so that they can choose the next training sample and then update their model based on the user's responses. The IME learning task is actually a case where the learning system (the IME) has access to a pool of unlabeled data (the multi-dimensional "truth set") and can ask an expert (the user) for the true label in a certain small number of instances. Furthermore, the performance of the learning system is assessed against the remaining training instances that are extracted from the database. In the learning literature this is called a pool-based, transductive active learning system[6]. These systems have been built using a variety of classifiers including SVM's.

Figure 22 shows the time required to train a user for a given training rate and number of training vectors. The goal was to limit the training time to the smallest possible duration, while maximizing the accuracy of the system (conflicting goals). For a presentation rate of one alert training sample every ten seconds (six alerts per minute), only 180 training samples can be presented in a 30 minute session. A reasonable training accuracy of one alert presentation error every 15 minutes (approximately a 1% error rate assuming six alerts per minute) will require the use of accurate historical reference models to initialize the classifier. A closely related issue should be addressed is how to detect and deal with inconsistent user input in an efficient and appropriate fashion.

[6] Tong, S. and D. Koller, "Support Vector Machine Active Learning with Applications to Text Classification," Journal of Machine Learning Research, November, 2001.

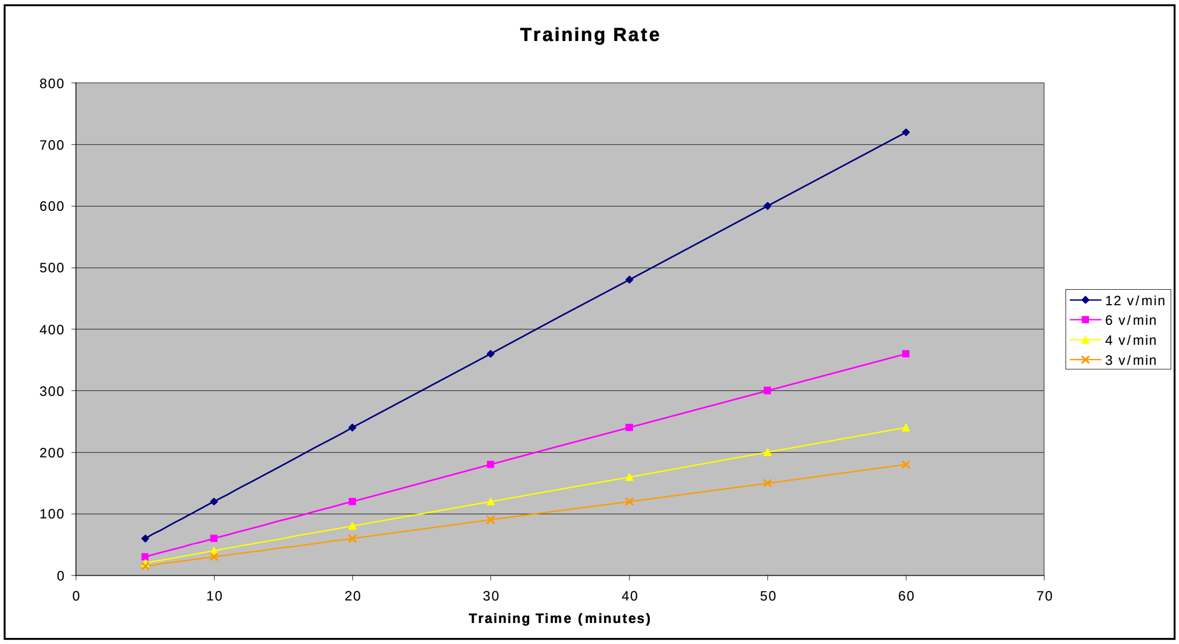


Figure 22: Training time for different rates and numbers of training samples.

6.3 Automated Alert Creation and Alert Presentation

In addition to the hardware-oriented aspects of TAMS described in Section 6.1, and information management oriented aspects described in Section 6.2, there are many software-oriented aspects of alert creation and alert presentation that should be considered by a follow on effort. Whether an alert is generated by a TAMS user, an INTEL source, or higher headquarters, this generation process addresses the alert payload. By this is meant that the text, audio, image, or video that is meant to convey the alert information. However, there are several meta-data fields that also have to be populated in an alert. While some of these fields are simple or mainly hardware dependent (e.g. alert time, alerting unit, soldier location), some, such as alert amplification and alert severity, must either be specified manually or by some automated process. If the person generating the alert is engaged in an operation, it would likely be difficult for him or her to specify this information to the system. We recommend that a follow on effort investigate methods for automating the alert creation process which would accept alert payload as one of the inputs.

A related issue that might be addressed is the validation of an alert. Depending on the user interface provided to generation an alert payload, it is possible that a user might inadvertently create an alert payload without intending to do so (e.g. snap an image of the ground). Being able to automatically validate an alert payload would reduce the probability of spurious alerts.

Alternatively, if a micro-electro mechanical system (MEMS) glove or similar technology is used as a user input device, some intelligent control algorithm could be built into the user input control software that could filter out spurious user input.

Alert presentation is an extremely important issue. It is essential to present alerts to the user in a way that does not confuse them and that minimizes the increase in his or her cognitive load. Ideally imagery and video would be preprocessed to enhance and possibly automatically annotate key features and to transform it to the user's point of view. Speech recognition algorithms should be applied to the audio portion of any alert to extract alert amplification metadata. Metadata information, perhaps relating the alert to a simplified map display clearly distinguishing relevant red and blue entities, must be presented to the user in an intuitive and useful format.

6.4 Operationally Focused Experimentation and Integration

The overwhelming issue encountered while developing and executing the experimentation campaign for this program was the difficulty in simulating an urban combat mission that was realistic and focused on an assessment of the increase in a soldier's operational capability with TAMS and not their computer skills.

If a follow on effort addressed the issue of alert creation in addition to alert presentation then some of our difficulties might be alleviated. However, Raytheon believes the choice of experimentation infrastructure ought to be reconsidered. In the following, four classes of alternatives are presented.

6.4.1 MOUT Area Based Experiments

The military maintains various MOUT areas for training purposes such as the McKenna MOUT Site at Fort Benning GA, or the MOUT areas at Fort Irwin CA, Quantico VA, Fort Polk LA, Fort Bragg NC or Camp Lejeune NC, to name a few. These areas contain physical buildings and other typical urban features and are often instrumented in some way or another for observation and evaluation purposes.

These areas also have a mechanism, typically the Miles II system, which allows soldiers to fire weapons at red forces and vise versa and for the system to "score" the hits. By this is meant that a laser gun and detector suit are used in combination to detect whether or not a shot fired from a laser gun hit an individual wearing a detector suit and where.

MOUT Area Based Experiments probably provide the most realistic mechanism for carrying out urban combat experiments but they require schedule coordination with the area administration staff. For the most part, it would be extremely difficult to simulate a blue or a red force, so human players will be needed to populate the entire scenario, which increases the cost of the experiment. While it may be too aggressive from a cost and time standpoint to run experiments at a MOUT site in the near term, it is worth investigating whether or not we could gain access to a MOUT site to create tactical alerts such as images and video.

6.4.2 Video Game Based Experiments

For the purposes of these experiments, we used a COTS video game to simulate urban combat missions. The video game used was very realistic, the 3D visualizations, while somewhat cartoon-ish, were extremely lifelike and the sound effects added much realism. The video game

also provided a fully automated and objective mechanism for collecting individual and team metrics. One major drawback is that the video game provided no hooks to support the integration of alert generation, scenario modifications, etc. The other major drawback is that the user had to be reasonably skilled with video games to be able to carry out the missions successfully. How well a user is able to aim at a hostile with a joy stick is not necessarily indicative of how well the same user is able to aim a rifle.

To overcome these drawbacks, we recommend developing a tailored urban combat simulation based on a video game platform so that it is possible to take advantage of the realism but can have more control over the user interface and over integration. Currently there is an open source effort to create a 3D gaming platform led by the same team of developers who created the America's Army video game. This is a very active effort. The gaming platform can be used as the basis of a TAMS experimentation simulation.

An alternative is Emergent Game Technology's Gamebryo platform. This platform has been used to support a Baghdad urban visualization at the JFETS conducted at Ft. Sill OK. Gamebryo is commercially available.

Taking this approach to experimentation would require software and scenario development and potentially tighter integration for data collection and alert generation than a physical MOUT site. The experiments could be stood up in any facility alleviating schedule dependencies.

6.4.3 Land-Based Simulation Experiments

The defense community at large has created a variety of land-based simulations which are extremely effective as training tools or course of action analysis tools from a command and control perspective. Unfortunately, we know of none that currently rival a video game for urban combat realism and squad level fidelity.

6.4.4 Hybrid Technology Based Experiments

Where current land-based simulations are being used most effectively to simulation urban combat operations is in combinations with other technologies. The JFETS at Ft. Sill OK is one example of where the three technologies previously listed have been combined to create a training facility for the U.S. Army. This facility consists of a physical mock sniper hideout (one room of a MOUT area as opposed to an entire area) with a simulated outdoor scene. The outdoor scene, developed with the Gamebryo platform, is displayed on liquid crystal display (LCD) monitors that are hung in the windows of the room. Currently there is no ability for the player to see into buildings or gain visual feedback from an engagement with a hostile. However, ongoing work is addressing these gaps.

We would not recommend developing such a facility but rather making arrangements to use an existing facility. The JFETS facility might be available to support experiments during off hours but we may have no ability to influence the vignette and the physical structure probably limits the missions with which we could experiment.

7. SYMBOLS, ABBREVIATIONS, AND ACRONYMS

1/A/1/75	1st Platoon, Company A, 1st Battalion, 75th Rangers - attached to 1st Armored Division with duty to assist the division with Stability and Support Operations (SASO) operations conducted in Baghdad and the immediate surrounding area.
1AD	1st Armored Division
1/35 Armor	1st Battalion, 35th Armor Regiment of the 2nd Brigade Combat Team (BCT). Armored Battalion organized as a Battalion Task Force equipped with M1A1 Abrams Tanks, M1A2 Bradley Fighting Vehicles and unarmored High-Mobility Multipurpose Wheeled Vehicles (HMMWVs). Main missions are to provide route security and patrols on major roads into and out of the "Green Zone".
2BCT	Second Brigade Combat Team
2BCT S-2	Intelligence officer for BCT
1/6 Infantry	1st Battalion, 6th Infantry Regiment of the 1st Brigade Combat Team (BCT). Mechanized Infantry Battalion organized as a Battalion Task Force equipped with M1A2 Bradleys, one company of Abrams Tanks and uparmored HMMWVs. Main missions are to provide support for SASO operations, man intersection checkpoints, and perform cordon and searches as required.
2/6 Infantry	2nd Battalion, 6th Infantry Regiment of the 1st Brigade Combat Team (BCT). Mechanized Infantry Battalion organized as a Battalion Task Force equipped with M1A2 Bradleys, one company of Abrams Tanks and uparmored HMMWVs. Main missions are to provide support for SASO operations, man intersection checkpoints, and perform cordon and searches as required.
3D	3 dimensional display of CERRTS
AAR	After Action Review
AFRL	Air Force Research Laboratory (controls all S&T funds in the USAF)
AGM	Alert Generation Module
Alert Severity Level	Symbol between S/1-S/5 specifying the severity of the information contained in an alert.
API	Application Program Interface
ASL	Alert Severity Level
AV or A/V	Audio/Video
B/1/6	Bravo company of the 1/6 infantry
BAA	Broad Agency Announcement
BFT	Blue Force Tracker
BLOB	Binary large object
BOS	Battlefield Operating System
C2	Command and Control
C2I	Command, Control, and Intelligence
CD	Compact Disc

CENTCOM	U.S. Central Command
CERRTS	Civil Emergency Reaction and Responder Training System
ConOps	Concept of Operations (plan by which equipment is used to achieve battle effects)
COTR	Contracting Officer's Technical Representative
COTS	Commercial-off-the-shelf
CP	Command Post
CPA	Coalition Provision Authority
DARPA	Defense Advanced Research Projects Agency
DCA	Data Collection and Analysis
DoD	Department of Defense
DOTMLPF	Doctrine, Organization, Training, Materiel, Leadership & Education, Personnel, and Facilities
DVD	Digital Versatile Disc
EO	Electro-optical
EOD	Explosives ordnance disposal
EPLRS	Enhanced Position Location Reporting System
FM	Frequency modulation
FOB	Forward Operating Base
FRD	Friendly
Gamebryo	A commercially available 3D graphics engine used in many commercial computer games
Ganges.com	A fictional web-site developed for the Simple Configuration
GMBSADS	Goggle-mounted binocular see-around (above/below) display system
GMD	Goggle-mounted display
Goggles	Synonym for GMBSADS
GPS	Global Positioning System
GUI	Graphical User Interface
HM	Head mounted
HMAS	Head Mounted Alerting System
HMBV	Head mounted binocular viewer (synonym for GMBSADS)
HMD	Head mounted display
HMMWVs	High-Mobility Multipurpose Wheeled Vehicles
HMV	Head mounted viewer
HQ	Headquarters
HUMINT	Human Intelligence
IED	Improvised explosive device
IEW	Intelligence and electronic warfare
IME	Information Management Engine
INTEL or Intel	Intelligence
INTSUM	Intelligence Summary
IP	Intellectual property or Internet Protocol
IR	Infra-red
JFETS	Joint Fires and Effects Training System
KIA	Killed in action

LCD	Liquid crystal display
LRIP	Low rate initial production
MAN	Metropolitan Area Network (e.g. IEEE 802.16; cell phone network)
MEMS	Micro-electro mechanical system
MIA	Missing in action
MNF-I	Multinational Forces Iraq
Mode	Text, audio, image, video
MOE	Measure of Effectiveness
MOP	Measure of Performance
MOS	Military occupational Specialty
MOUT	Military Operations on Urban Terrain
MSR	Main supply route
Multi-modal alert	Alerts were presented to user with both audio and video stimuli.
OBJ	Objective
ODA	Operational Detachment Alpha
ODA Team	The lowest level unit in the Special Forces
ONS	Operational Needs Statement
OPCON	Operational Control
OPS	Operations
OR	Operating room
PA	Peripheral awareness (SA plus stimulation of the user with alerts that present the user with information he or she cannot anticipate)
Payload	A term frequently used to refer to the actual content of a message in a message stream to discriminate it from the metadata parameters often contained in the message. For TAMS alerts, payload refers to the text string, audio clip, image, or video clip upon which the alert is based.
PC	Personal computer
PSDS2	Persistent Surveillance and Dissemination System of Systems
ROE	Rules of Engagement
s	seconds
S-2	Intelligence Officer
SA	Situational Awareness (user defines the display based on his or her interests)
SAL	Soldier Activity Level
SAM	Surface to air missile
SASO	Stability and Support Operations
Scenario	A collection of configuration files and databases needed to create a simulation of a vignette in CERRTS
Server	
SF	Special Forces (Army)
SGI	Silicon Graphics Infinity (computer)
Smart Book	A quick reference users guide put together to prepare a user for an exercise or experiment.
SME	Subject Matter Expert

Smoketest	A test where all the major modules of a software system are integrated together and a simple end to end test is run to make sure that the interfaces are well understood and working.
SOC	Surveillance Operation Center
SOF	Special Operations Force
Soldier Activity Level	Number from 1-5 specifying how busy the soldier is at this instant in time
SoW	Statement of Work
SVGA	Super Video Graphic Array
SVM	Support vector machine
TAMS	Tactical Alert Management System
TAMS Support Team	Two Raytheon and one HRL employee who were present at and supported the experiments.
TRADOC	U.S. Army Training and Doctrine Command
TTP	Tactics Techniques and Procedures
TTS	Text to Speech – software that synthesizes a voice from a text input string.
TOC	Tactical Operations Center
TUSK	Tank Urban Survivability Kit
TWS	Thermal Weapon Site
UAV	Unmanned air vehicle
UDOP	User defined operating picture
URL	Universal Resource Locator
US	United States
USAF	United States Air Force
Vignette	A description of an operational event or sequence of events developed for the purpose of testing or experimenting with ConOps
VBIED	Vehicle-borne IED
WIA	Wounded in action

APPENDIX A.
INTELLECTUAL PROPERTY RESULTING FROM THIS PROGRAM

A1. Raytheon Company

Title: "TAMS ConOps"
Inventors: Susan Gottschlich, Robert Gray

A2. HRL Laboratories LLC

Title: "IME training algorithm"
Inventors: Michael Daily, Ron Azuma, Youngkwan Cho

A3. Joint

Title: "Alert interface ontology"
Inventors: Michael Daily, Susan Gottschlich, Robert Gray, Ron Azuma

APPENDIX B.
PUBLICATIONS AND PRESENTATIONS

B1. Publications

Mike Daily, Ron Azuma, Youngkwan Cho, Troy Rockwood, and Susan Gottschlich, "Tactical Alert Management," submitted for publication to *International Conference on Artificial Intelligence*, 2006.

B2. Presentations

Mike Daily, Ron Azuma, Youngkwan Cho, Troy Rockwood, and Susan Gottschlich, "Tactical Alert Management," submitted for presentation to *International Conference on Artificial Intelligence*, 2006.

APPENDIX C.
PROFESSIONAL PERSONNEL ASSOCIATED WITH THIS PROGRAM

Persons who contributed to technical work on this effort by role are listed below.

Company	Name	Role
Raytheon Company		
	Dr. Susan Gottschlich	Principal Investigator and TAMS Support Team, Engineering Fellow
	Mr. John Baird	Raytheon C2 Integrated Systems, Director
	Mr. Robert Gray	Military Ops SME and TAMS Support Team, Systems Engineer
	Mr. Craig Parkhill	CERRTS Support, Systems Engineer
	Dr. Gabriel Pei	Consultant for ConOps, Program Manager
	Mr. Michael Blose	Military Ops SME, Systems Engineer
	MG (Ret) Dean Cash	Consultant for Experiment Definition, Director
	MSGT (Ret) Wilbur Adams	Consultant for Experiments, Manager
HRL Laboratories LLC		
	Mr. Michael Daily	Co-Principal Investigator, Program Manager
	Dr. Ron Azuma	Augmented Realty SME, IME Architect
	Mr. Youngkwan Cho	IME Software Development and TAMS Support Team, Software Engineer